IMDB MOVIE ANALYSIS

-Saksham Gupta

[71]:
```
# Filtering out the warnings

import warnings

warnings.filterwarnings('ignore')
```

[72]:
```
# Importing the required libraries

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

1 IMDb Movie Assignment

You have the data for the 100 top-rated movies from the past decade along with various pieces of information about the movie, its actors, and the voters who have rated these movies online. In this assignment, you will try to find some interesting insights into these movies and their voters, using Python.

1.1 Task 1: Reading the data

-

1.1.1 Subtask 1.1: Read the Movies Data.

Read the movies data file provided and store it in a dataframe `movies`.

[73]:
```
# Read the csv file using 'read_csv'. Please write your dataset location here.

movies= pd.read_csv(r"D:\Courses\Upgrad\PYTHON\DATA TOOLKIT\IMDB␣
↪Assignment\Movie+Assignment+Data.csv")
pd.set_option('display.max_columns', 70)
```

-

1.1.2 Subtask 1.2: Inspect the Dataframe

Inspect the dataframe for dimensions, null-values, and summary of different numeric columns.

[74]:
```
#Check the number of rows and columns in the dataframe

movies.shape
```

[74]:
```
(100, 62)
```

[75]:
```
# Check the column-wise info of the dataframe

movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 62 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Title                   100 non-null    object
 1   title_year              100 non-null    int64
 2   budget                  100 non-null    int64
 3   Gross                   100 non-null    int64
 4   actor_1_name            100 non-null    object
 5   actor_2_name            100 non-null    object
 6   actor_3_name            100 non-null    object
 7   actor_1_facebook_likes  100 non-null    int64
 8   actor_2_facebook_likes  99 non-null     float64
 9   actor_3_facebook_likes  98 non-null     float64
 10  IMDb_rating             100 non-null    float64
 11  genre_1                 100 non-null    object
 12  genre_2                 97 non-null     object
 13  genre_3                 74 non-null     object
 14  MetaCritic              95 non-null     float64
 15  Runtime                 100 non-null    int64
 16  CVotes10                100 non-null    int64
 17  CVotes09                100 non-null    int64
 18  CVotes08                100 non-null    int64
 19  CVotes07                100 non-null    int64
 20  CVotes06                100 non-null    int64
 21  CVotes05                100 non-null    int64
 22  CVotes04                100 non-null    int64
 23  CVotes03                100 non-null    int64
 24  CVotes02                100 non-null    int64
 25  CVotes01                100 non-null    int64
 26  CVotesMale              100 non-null    int64
 27  CVotesFemale            100 non-null    int64
 28  CVotesU18               100 non-null    int64
```

```
 29   CVotesU18M             100 non-null    int64
 30   CVotesU18F             100 non-null    int64
 31   CVotes1829             100 non-null    int64
 32   CVotes1829M            100 non-null    int64
 33   CVotes1829F            100 non-null    int64
 34   CVotes3044             100 non-null    int64
 35   CVotes3044M            100 non-null    int64
 36   CVotes3044F            100 non-null    int64
 37   CVotes45A              100 non-null    int64
 38   CVotes45AM             100 non-null    int64
 39   CVotes45AF             100 non-null    int64
 40   CVotes1000             100 non-null    int64
 41   CVotesUS               100 non-null    int64
 42   CVotesnUS              100 non-null    int64
 43   VotesM                 100 non-null    float64
 44   VotesF                 100 non-null    float64
 45   VotesU18               100 non-null    float64
 46   VotesU18M              100 non-null    float64
 47   VotesU18F              100 non-null    float64
 48   Votes1829              100 non-null    float64
 49   Votes1829M             100 non-null    float64
 50   Votes1829F             100 non-null    float64
 51   Votes3044              100 non-null    float64
 52   Votes3044M             100 non-null    float64
 53   Votes3044F             100 non-null    float64
 54   Votes45A               100 non-null    float64
 55   Votes45AM              100 non-null    float64
 56   Votes45AF              100 non-null    float64
 57   Votes1000              100 non-null    float64
 58   VotesUS                100 non-null    float64
 59   VotesnUS               100 non-null    float64
 60   content_rating         100 non-null    object
 61   Country                100 non-null    object
dtypes: float64(21), int64(32), object(9)
memory usage: 48.6+ KB
```

[76]:
```
# Check the summary for the numeric columns
movies.describe()
```

[76]:
```
        title_year        budget         Gross  actor_1_facebook_likes  \
count   100.000000  1.000000e+02  1.000000e+02              100.000000
mean   2012.820000  7.838400e+07  1.468679e+08            13407.270000
std       1.919491  7.445295e+07  1.454004e+08            10649.037862
min    2010.000000  3.000000e+06  2.238380e+05               39.000000
25%    2011.000000  1.575000e+07  4.199752e+07             1000.000000
50%    2013.000000  4.225000e+07  1.070266e+08            13000.000000
75%    2014.000000  1.500000e+08  2.107548e+08            20000.000000
```

max	2016.000000	2.600000e+08	9.366622e+08	35000.000000

	actor_2_facebook_likes	actor_3_facebook_likes	IMDb_rating	\
count	99.000000	98.000000	100.000000	
mean	7377.303030	3002.153061	7.883000	
std	13471.568216	6940.301133	0.247433	
min	12.000000	0.000000	7.500000	
25%	580.000000	319.750000	7.700000	
50%	1000.000000	626.500000	7.800000	
75%	11000.000000	1000.000000	8.100000	
max	96000.000000	46000.000000	8.800000	

	MetaCritic	Runtime	CVotes10	CVotes09	CVotes08	\
count	95.000000	100.000000	100.000000	100.000000	100.000000	
mean	78.252632	126.420000	73212.160000	92404.170000	125762.230000	
std	9.122066	19.050799	82669.594746	75666.918775	62162.752481	
min	62.000000	91.000000	6420.000000	7321.000000	11668.000000	
25%	72.000000	114.750000	30587.000000	47098.250000	83207.500000	
50%	78.000000	124.000000	54900.500000	71376.000000	117405.500000	
75%	83.500000	136.250000	80639.000000	115240.250000	166742.500000	
max	100.000000	180.000000	584839.000000	485218.000000	304457.000000	

	CVotes07	CVotes06	CVotes05	CVotes04	CVotes03	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	76715.280000	27841.450000	10171.750000	4330.110000	2477.070000	
std	32586.358624	12883.912563	5480.914204	2713.608902	1768.167506	
min	8558.000000	3370.000000	1162.000000	456.000000	227.000000	
25%	54934.500000	17834.500000	6026.000000	2293.500000	1172.750000	
50%	76862.000000	27072.000000	9123.000000	3801.000000	1950.000000	
75%	100789.000000	37561.750000	13008.500000	5571.000000	3309.500000	
max	162604.000000	67579.000000	27957.000000	12286.000000	7868.000000	

	CVotes02	CVotes01	CVotesMale	CVotesFemale	CVotesU18	\
count	100.000000	100.000000	1.000000e+02	100.000000	100.000000	
mean	1711.570000	4084.350000	2.714462e+05	64468.860000	1769.050000	
std	1318.631164	3489.778403	1.676580e+05	39117.954828	1324.522818	
min	158.000000	293.000000	2.244100e+04	9552.000000	121.000000	
25%	761.750000	1532.750000	1.719362e+05	35188.750000	756.250000	
50%	1320.000000	2933.000000	2.381995e+05	58539.500000	1420.000000	
75%	2334.500000	5296.000000	3.383265e+05	82355.000000	2478.750000	
max	5751.000000	15768.000000	1.044318e+06	239796.000000	5735.000000	

	CVotesU18M	CVotesU18F	CVotes1829	CVotes1829M	CVotes1829F	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	1318.950000	435.660000	165863.770000	128473.140000	35739.710000	
std	1040.959361	366.499996	103522.932356	83188.725257	23092.041366	
min	95.000000	20.000000	15959.000000	10150.000000	4370.000000	

```
25%      579.000000   159.750000  109797.750000   77537.750000   17637.750000
50%      998.000000   325.500000  139321.500000  111362.000000   32639.500000
75%     1790.750000   660.000000  212788.250000  160442.000000   47301.000000
max     4596.000000  1910.000000  655187.000000  512411.000000  136770.000000

          CVotes3044    CVotes3044M   CVotes3044F    CVotes45A    CVotes45AM  \
count     100.000000     100.000000    100.000000   100.000000    100.000000
mean   124561.100000  103437.920000  19371.540000  24170.160000  19653.260000
std     72089.245359   61116.374727  11324.680698  12841.696513  10669.041306
min     12174.000000    9280.000000   2682.000000   1899.000000   1496.000000
25%     78866.250000   64930.250000  12030.500000  15288.750000  12154.750000
50%    111317.500000   94010.500000  17763.000000  21869.500000  17782.000000
75%    158760.750000  135668.750000  24859.500000  30043.000000  24780.000000
max    472680.000000  392845.000000  73555.000000  79634.000000  65508.000000

         CVotes45AF  CVotes1000        CVotesUS       CVotesnUS      VotesM  \
count    100.000000  100.000000      100.000000      100.000000  100.000000
mean    4093.530000  624.370000    54096.740000   188245.680000    7.852000
std     2187.955655  133.244262    32115.288162   106902.256347    0.263036
min      355.000000  198.000000     3678.000000    19009.000000    7.400000
25%     2492.750000  545.000000    33328.750000   121863.500000    7.675000
50%     3771.500000  639.500000    48560.000000   169132.000000    7.800000
75%     5342.250000  727.000000    70167.000000   239628.250000    8.000000
max    12795.000000  885.000000   212524.000000   707266.000000    8.800000

           VotesF    VotesU18   VotesU18M   VotesU18F   Votes1829  Votes1829M  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     7.910000    8.205000    8.174000    8.215000    8.005000    8.005000
std      0.298312    0.305959    0.358059    0.381484    0.271686    0.281904
min      7.300000    7.500000    7.400000    7.200000    7.600000    7.600000
25%      7.700000    8.000000    7.900000    7.975000    7.800000    7.800000
50%      7.900000    8.200000    8.150000    8.300000    8.000000    8.000000
75%      8.100000    8.400000    8.400000    8.500000    8.125000    8.100000
max      8.700000    9.100000    9.100000    9.000000    9.000000    9.000000

       Votes1829F   Votes3044  Votes3044M  Votes3044F   Votes45A  Votes45AM  \
count  100.000000  100.000000  100.000000  100.000000  100.00000 100.000000
mean     7.982000    7.732000    7.723000    7.780000    7.65100   7.624000
std      0.321417    0.251814    0.260479    0.282128    0.21485   0.213258
min      7.300000    7.300000    7.200000    7.200000    7.10000   7.100000
25%      7.700000    7.600000    7.500000    7.600000    7.50000   7.475000
50%      8.000000    7.700000    7.700000    7.800000    7.65000   7.600000
75%      8.200000    7.900000    7.900000    8.000000    7.80000   7.800000
max      8.800000    8.700000    8.700000    8.500000    8.10000   8.100000

        Votes45AF   Votes1000     VotesUS    VotesnUS
count  100.000000  100.000000  100.000000  100.000000
```

```
mean     7.770000    7.274000    7.958000    7.793000
std      0.301344    0.361987    0.232327    0.264099
min      7.000000    6.400000    7.500000    7.300000
25%      7.500000    7.100000    7.800000    7.600000
50%      7.800000    7.300000    7.950000    7.750000
75%      7.925000    7.500000    8.100000    7.925000
max      8.500000    8.200000    8.700000    8.800000
```

1.2 Task 2: Data Analysis

Now that we have loaded the dataset and inspected it, we see that most of the data is in place. As of now, no data cleaning is required, so let's start with some data manipulation, analysis, and visualisation to get various insights about the data.

-

1.2.1 Subtask 2.1: Reduce those Digits!

These numbers in the `budget` and `gross` are too big, compromising its readability. Let's convert the unit of the `budget` and `gross` columns from `$` to `million $` first.

[77]:
```
movies.head()
```

[77]:
```
                   Title  title_year     budget      Gross       actor_1_name  \
0              La La Land        2016   30000000  151101803        Ryan Gosling
1                Zootopia        2016  150000000  341268248  Ginnifer Goodwin
2                    Lion        2016   12000000   51738905         Dev Patel
3                 Arrival        2016   47000000  100546139         Amy Adams
4  Manchester by the Sea        2016    9000000   47695371     Casey Affleck

        actor_2_name     actor_3_name  actor_1_facebook_likes  \
0        Emma Stone       Amiée Conn                   14000
1     Jason Bateman       Idris Elba                    2800
2      Nicole Kidman      Rooney Mara                   33000
3      Jeremy Renner  Forest Whitaker                   35000
4  Michelle Williams     Kyle Chandler                     518

   actor_2_facebook_likes  actor_3_facebook_likes  IMDb_rating    genre_1  \
0                 19000.0                     NaN          8.2     Comedy
1                 28000.0                 27000.0          8.1  Animation
2                 96000.0                  9800.0          8.1  Biography
3                  5300.0                     NaN          8.0      Drama
4                 71000.0                  3300.0          7.9      Drama

     genre_2 genre_3  MetaCritic  Runtime  CVotes10  CVotes09  CVotes08  \
0      Drama   Music        93.0      128     74245     71191     64640
1  Adventure  Comedy        78.0      108     53626     70912    102352
2      Drama     NaN        69.0      118     23325     29830     40564
```

```
3   Mystery  Sci-Fi        81.0       116      55533      87850     109536
4       NaN     NaN        96.0       137      18191      33532      46596

   CVotes07  CVotes06  CVotes05  CVotes04  CVotes03  CVotes02  CVotes01  \
0     38831     17377      8044      3998      2839      2407      6802
1     57261     16719      4539      1467       733       496      1386
2     20296      5842      1669       558       309       182       493
3     65440     26913     10556      5057      3083      2194      4734
4     29626     11879      4539      1976      1233       888      1834

   CVotesMale  CVotesFemale  CVotesU18  CVotesU18M  CVotesU18F  CVotes1829  \
0      157693         56713       2675        1784         868      113008
1      176202         52345       2362        1641         706      119637
2       68921         24977        702         477         220       42962
3      237437         46272       1943        1544         376      126301
4       92452         22834        855         681         166       55475

   CVotes1829M  CVotes1829F  CVotes3044  CVotes3044M  CVotes3044F  CVotes45A  \
0        78998        32730       66058        50835        14165      15765
1        87499        30813       75474        61358        13034      12353
2        29729        12780       34297        26384         7413       9054
3       101741        23163      111985        95005        15227      24027
4        43467        11378       40645        32983         7053      11361

   CVotes45AM  CVotes45AF  CVotes1000  CVotesUS  CVotesnUS  VotesM  VotesF  \
0       12148        3302         454     33360     117987     8.2     8.1
1        9959        2151         518     35975     122844     8.0     8.3
2        6714        2184         298     13478      53931     8.0     8.4
3       20118        3440         537     42062     163774     7.9     8.0
4        8862        2306         402     20287      65837     7.9     7.7

   VotesU18  VotesU18M  VotesU18F  Votes1829  Votes1829M  Votes1829F  \
0       8.9        9.0        8.7        8.4         8.4         8.2
1       8.4        8.3        8.7        8.2         8.1         8.4
2       8.3        8.2        8.7        8.1         8.0         8.4
3       8.6        8.6        8.4        8.2         8.2         8.1
4       8.5        8.5        8.1        8.0         8.1         7.8

   Votes3044  Votes3044M  Votes3044F  Votes45A  Votes45AM  Votes45AF  \
0        7.9         7.9         7.8       7.6        7.6        7.5
1        7.8         7.8         8.1       7.8        7.8        8.1
2        8.0         7.9         8.2       8.0        7.9        8.4
3        7.8         7.8         7.8       7.6        7.6        7.7
4        7.7         7.7         7.7       7.6        7.6        7.6

   Votes1000  VotesUS  VotesnUS content_rating    Country
0        7.1      8.3       8.1          PG-13        USA
```

```
1       7.6      8.0       8.0            PG        USA
2       7.1      8.1       8.0         PG-13  Australia
3       7.3      8.0       7.9         PG-13        USA
4       7.1      7.9       7.8             R        USA
```

[78]:

```
# Divide the 'gross' and 'budget' columns by 1000000 to convert '$' to 'million␣
↪$'
movies['budget'] = movies['budget']/1000000
movies['Gross'] = movies['Gross']/1000000
```

[79]:

```
movies.head()
```

[79]:

```
                    Title  title_year  budget       Gross      actor_1_name  \
0               La La Land        2016    30.0  151.101803      Ryan Gosling
1                 Zootopia        2016   150.0  341.268248  Ginnifer Goodwin
2                     Lion        2016    12.0   51.738905         Dev Patel
3                  Arrival        2016    47.0  100.546139        Amy Adams
4  Manchester by the Sea        2016     9.0   47.695371     Casey Affleck

        actor_2_name      actor_3_name  actor_1_facebook_likes  \
0        Emma Stone        Amiée Conn                   14000
1     Jason Bateman        Idris Elba                    2800
2     Nicole Kidman       Rooney Mara                   33000
3     Jeremy Renner  Forest Whitaker                    35000
4  Michelle Williams     Kyle Chandler                     518

   actor_2_facebook_likes  actor_3_facebook_likes  IMDb_rating    genre_1  \
0                 19000.0                     NaN          8.2     Comedy
1                 28000.0                 27000.0          8.1  Animation
2                 96000.0                  9800.0          8.1  Biography
3                  5300.0                     NaN          8.0      Drama
4                 71000.0                  3300.0          7.9      Drama

     genre_2 genre_3  MetaCritic  Runtime  CVotes10  CVotes09  CVotes08  \
0      Drama   Music        93.0      128     74245     71191     64640
1  Adventure  Comedy        78.0      108     53626     70912    102352
2      Drama     NaN        69.0      118     23325     29830     40564
3    Mystery  Sci-Fi        81.0      116     55533     87850    109536
4        NaN     NaN        96.0      137     18191     33532     46596

   CVotes07  CVotes06  CVotes05  CVotes04  CVotes03  CVotes02  CVotes01  \
0     38831     17377      8044      3998      2839      2407      6802
1     57261     16719      4539      1467       733       496      1386
2     20296      5842      1669       558       309       182       493
3     65440     26913     10556      5057      3083      2194      4734
4     29626     11879      4539      1976      1233       888      1834
```

```
   CVotesMale  CVotesFemale  CVotesU18  CVotesU18M  CVotesU18F  CVotes1829  \
0      157693         56713       2675        1784         868      113008   
1      176202         52345       2362        1641         706      119637   
2       68921         24977        702         477         220       42962   
3      237437         46272       1943        1544         376      126301   
4       92452         22834        855         681         166       55475   

   CVotes1829M  CVotes1829F  CVotes3044  CVotes3044M  CVotes3044F  CVotes45A  \
0        78998        32730       66058        50835        14165      15765   
1        87499        30813       75474        61358        13034      12353   
2        29729        12780       34297        26384         7413       9054   
3       101741        23163      111985        95005        15227      24027   
4        43467        11378       40645        32983         7053      11361   

   CVotes45AM  CVotes45AF  CVotes1000  CVotesUS  CVotesnUS  VotesM  VotesF  \
0       12148        3302         454     33360     117987     8.2     8.1   
1        9959        2151         518     35975     122844     8.0     8.3   
2        6714        2184         298     13478      53931     8.0     8.4   
3       20118        3440         537     42062     163774     7.9     8.0   
4        8862        2306         402     20287      65837     7.9     7.7   

   VotesU18  VotesU18M  VotesU18F  Votes1829  Votes1829M  Votes1829F  \
0       8.9        9.0        8.7        8.4         8.4         8.2   
1       8.4        8.3        8.7        8.2         8.1         8.4   
2       8.3        8.2        8.7        8.1         8.0         8.4   
3       8.6        8.6        8.4        8.2         8.2         8.1   
4       8.5        8.5        8.1        8.0         8.1         7.8   

   Votes3044  Votes3044M  Votes3044F  Votes45A  Votes45AM  Votes45AF  \
0        7.9         7.9         7.8       7.6        7.6        7.5   
1        7.8         7.8         8.1       7.8        7.8        8.1   
2        8.0         7.9         8.2       8.0        7.9        8.4   
3        7.8         7.8         7.8       7.6        7.6        7.7   
4        7.7         7.7         7.7       7.6        7.6        7.6   

   Votes1000  VotesUS  VotesnUS content_rating    Country  
0        7.1      8.3       8.1          PG-13        USA  
1        7.6      8.0       8.0             PG        USA  
2        7.1      8.1       8.0          PG-13  Australia  
3        7.3      8.0       7.9          PG-13        USA  
4        7.1      7.9       7.8              R        USA  
```

-

1.2.2 Subtask 2.2: Let's Talk Profit!

1. Create a new column called `profit` which contains the difference of the two columns: `gross` and `budget`.

2. Sort the dataframe using the `profit` column as reference.
3. Extract the top ten profiting movies in descending order and store them in a new dataframe - `top10`.
4. Plot a scatter or a joint plot between the columns `budget` and `profit` and write a few words on what you observed.
5. Extract the movies with a negative profit and store them in a new dataframe - `neg_profit`

[81]:
```python
# Create the new column named 'profit' by subtracting the 'budget' column from
↪the 'gross' column

movies['Profit'] = movies['Gross'] - movies['budget']
movies.head()
```

[81]:
```
                   Title  title_year  budget       Gross      actor_1_name  \
0              La La Land        2016    30.0  151.101803      Ryan Gosling
1                Zootopia        2016   150.0  341.268248  Ginnifer Goodwin
2                    Lion        2016    12.0   51.738905         Dev Patel
3                 Arrival        2016    47.0  100.546139         Amy Adams
4  Manchester by the Sea        2016     9.0   47.695371     Casey Affleck

        actor_2_name     actor_3_name  actor_1_facebook_likes  \
0        Emma Stone       Amiée Conn                   14000
1      Jason Bateman       Idris Elba                    2800
2      Nicole Kidman      Rooney Mara                   33000
3      Jeremy Renner  Forest Whitaker                   35000
4  Michelle Williams    Kyle Chandler                     518

   actor_2_facebook_likes  actor_3_facebook_likes  IMDb_rating    genre_1  \
0                 19000.0                     NaN          8.2     Comedy
1                 28000.0                 27000.0          8.1  Animation
2                 96000.0                  9800.0          8.1  Biography
3                  5300.0                     NaN          8.0      Drama
4                 71000.0                  3300.0          7.9      Drama

     genre_2 genre_3  MetaCritic  Runtime  CVotes10  CVotes09  CVotes08  \
0      Drama   Music        93.0      128     74245     71191     64640
1  Adventure  Comedy        78.0      108     53626     70912    102352
2      Drama     NaN        69.0      118     23325     29830     40564
3    Mystery  Sci-Fi        81.0      116     55533     87850    109536
4        NaN     NaN        96.0      137     18191     33532     46596

   CVotes07  CVotes06  CVotes05  CVotes04  CVotes03  CVotes02  CVotes01  \
0     38831     17377      8044      3998      2839      2407      6802
1     57261     16719      4539      1467       733       496      1386
2     20296      5842      1669       558       309       182       493
3     65440     26913     10556      5057      3083      2194      4734
```

```
4     29626     11879      4539      1976      1233       888      1834

   CVotesMale  CVotesFemale  CVotesU18  CVotesU18M  CVotesU18F  CVotes1829  \
0      157693         56713       2675        1784         868      113008
1      176202         52345       2362        1641         706      119637
2       68921         24977        702         477         220       42962
3      237437         46272       1943        1544         376      126301
4       92452         22834        855         681         166       55475

   CVotes1829M  CVotes1829F  CVotes3044  CVotes3044M  CVotes3044F  CVotes45A  \
0        78998        32730       66058        50835        14165      15765
1        87499        30813       75474        61358        13034      12353
2        29729        12780       34297        26384         7413       9054
3       101741        23163      111985        95005        15227      24027
4        43467        11378       40645        32983         7053      11361

   CVotes45AM  CVotes45AF  CVotes1000  CVotesUS  CVotesnUS  VotesM  VotesF  \
0       12148        3302         454     33360     117987     8.2     8.1
1        9959        2151         518     35975     122844     8.0     8.3
2        6714        2184         298     13478      53931     8.0     8.4
3       20118        3440         537     42062     163774     7.9     8.0
4        8862        2306         402     20287      65837     7.9     7.7

   VotesU18  VotesU18M  VotesU18F  Votes1829  Votes1829M  Votes1829F  \
0       8.9        9.0        8.7        8.4         8.4         8.2
1       8.4        8.3        8.7        8.2         8.1         8.4
2       8.3        8.2        8.7        8.1         8.0         8.4
3       8.6        8.6        8.4        8.2         8.2         8.1
4       8.5        8.5        8.1        8.0         8.1         7.8

   Votes3044  Votes3044M  Votes3044F  Votes45A  Votes45AM  Votes45AF  \
0        7.9         7.9         7.8       7.6        7.6        7.5
1        7.8         7.8         8.1       7.8        7.8        8.1
2        8.0         7.9         8.2       8.0        7.9        8.4
3        7.8         7.8         7.8       7.6        7.6        7.7
4        7.7         7.7         7.7       7.6        7.6        7.6

   Votes1000  VotesUS  VotesnUS content_rating    Country      Profit
0        7.1      8.3       8.1          PG-13        USA  121.101803
1        7.6      8.0       8.0             PG        USA  191.268248
2        7.1      8.1       8.0          PG-13  Australia   39.738905
3        7.3      8.0       7.9          PG-13        USA   53.546139
4        7.1      7.9       7.8              R        USA   38.695371
```

[82]:

```
# Sort the dataframe with the 'profit' column as reference using the
↪'sort_values' function. Make sure to set the argument
#'ascending' to 'False'
```

```python
movies = movies.sort_values(by='Profit' , ascending=False)
```

[83]:
```python
movies.head()
```

[83]:
```
                                        Title  title_year  budget  \
97  Star Wars: Episode VII - The Force Awakens        2015   245.0
11                               The Avengers        2012   220.0
47                                   Deadpool        2016    58.0
32               The Hunger Games: Catching Fire        2013   130.0
12                                Toy Story 3        2010   200.0

         Gross       actor_1_name       actor_2_name           actor_3_name  \
97  936.662225        Doug Walker         Rob Walker                      0
11  623.279547    Chris Hemsworth  Robert Downey Jr.      Scarlett Johansson
47  363.024263      Ryan Reynolds          Ed Skrein         Stefan Kapicic
32  424.645577  Jennifer Lawrence     Josh Hutcherson  Sandra Ellis Lafferty
12  414.984497          Tom Hanks  John Ratzenberger            Don Rickles

    actor_1_facebook_likes  actor_2_facebook_likes  actor_3_facebook_likes  \
97                     131                    12.0                     0.0
11                   26000                 21000.0                 19000.0
47                   16000                   805.0                   361.0
32                   34000                 14000.0                   523.0
12                   15000                  1000.0                   721.0

    IMDb_rating    genre_1    genre_2  genre_3  MetaCritic  Runtime  CVotes10  \
97          8.1     Action  Adventure  Fantasy        81.0      136    155391
11          8.1     Action     Sci-Fi      NaN        69.0      143    260257
47          8.0     Action  Adventure   Comedy        65.0      108    147467
32          7.6     Action  Adventure  Mystery        76.0      146     85219
12          8.3  Animation  Adventure   Comedy        92.0      103    139773

    CVotes09  CVotes08  CVotes07  CVotes06  CVotes05  CVotes04  CVotes03  \
97    161810    166378     99402     40734     18060      8751      5970
11    234203    264290    162604     67579     27957     12176      7201
47    147966    170810    105717     41811     15510      7046      4273
32     83874    150153    121748     50575     18571      7591      4094
12    149992    158704     88289     31291     11850      4859      2932

    CVotes02  CVotes01  CVotesMale  CVotesFemale  CVotesU18  CVotesU18M  \
97      4489     15768      425971         68664       4722        3919
11      4996     15528      691783        151617       4953        3767
47      3037      8538      391955         79804       4598        3601
32      2675      6978      307237        115421       3650        1956
12      2119      6586      389014         98386       3202        2405
```

```
    CVotesU18F  CVotes1829  CVotes1829M  CVotes1829F  CVotes3044  CVotes3044M  \
97         768      220467       183671        34366      187138       162918
11        1150      432999       343012        85465      295318       247617
47         969      232840       186139        44316      159222       135428
32        1664      218884       148652        67934      140683       109976
12         776      260519       199962        58366      169886       140253

    CVotes3044F  CVotes45A  CVotes45AM  CVotes45AF  CVotes1000  CVotesUS  \
97        21362      42942       36441        5729         712     85141
11        43303      54282       44183        9138         842    145826
47        21521      28753       24218        4009         667     67933
32        28735      27789       21545        5771         693     68521
12        27658      32457       26171        5806         769    105490

    CVotesnUS  VotesM  VotesF  VotesU18  VotesU18M  VotesU18F  Votes1829  \
97     250769     8.0     8.3       8.5        8.5        8.6        8.2
11     423958     8.0     8.2       8.2        8.2        8.5        8.1
47     241138     8.0     8.1       8.4        8.4        8.6        8.1
32     221430     7.4     8.1       8.0        7.7        8.5        7.8
12     267692     8.3     8.3       8.2        8.3        8.0        8.4

    Votes1829M  Votes1829F  Votes3044  Votes3044M  Votes3044F  Votes45A  \
97         8.2         8.3        8.0         7.9         8.2       7.9
11         8.1         8.3        8.0         8.0         8.1       7.9
47         8.1         8.1        7.9         7.9         7.9       7.8
32         7.6         8.2        7.3         7.2         7.9       7.3
12         8.5         8.4        8.2         8.2         8.3       8.1

    Votes45AM  Votes45AF  Votes1000  VotesUS  VotesnUS content_rating Country  \
97        7.8        8.2        7.7      8.2       7.9          PG-13     USA
11        7.9        8.1        7.4      8.3       7.9          PG-13     USA
47        7.8        7.9        7.3      8.1       7.9              R     USA
32        7.2        7.9        6.7      7.7       7.4          PG-13     USA
12        8.1        8.1        8.1      8.5       8.3              G     USA

        Profit
97  691.662225
11  403.279547
47  305.024263
32  294.645577
12  214.984497
```

[84]:

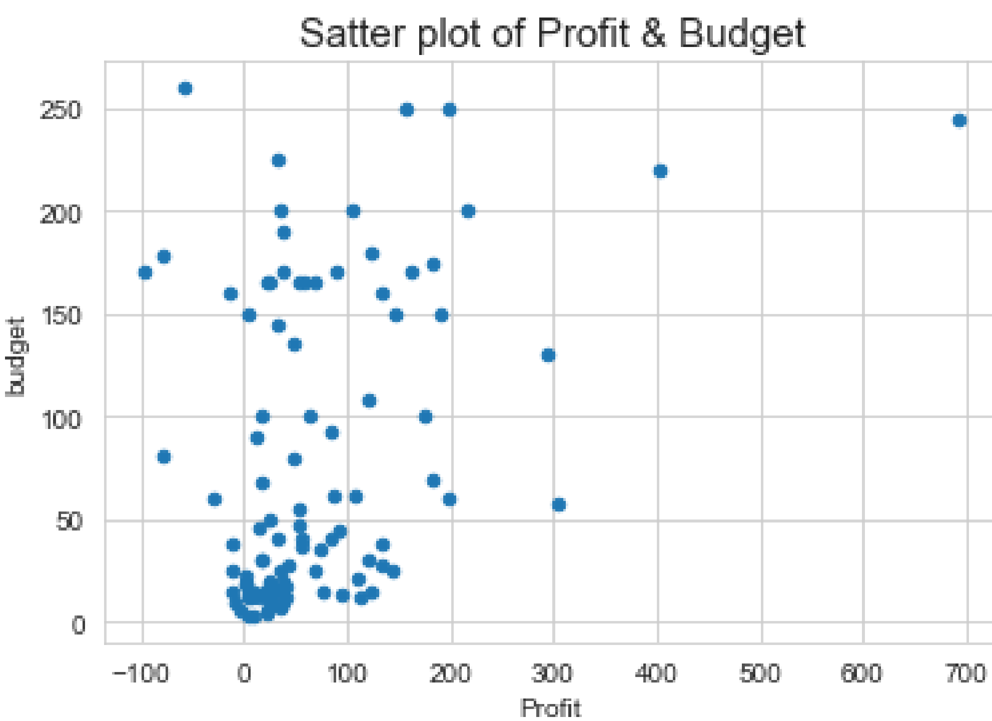
```
# Get the top 10 profitable movies by using position based indexing. Specify
↪the rows till 10 (0-9)

top_10 = movies['Title'].iloc[:10]
```

[85]:
```
top_10
```

[85]:
```
97     Star Wars: Episode VII - The Force Awakens
11                                   The Avengers
47                                       Deadpool
32                 The Hunger Games: Catching Fire
12                                    Toy Story 3
8                           The Dark Knight Rises
45                                 The Lego Movie
1                                        Zootopia
41                                   Despicable Me
18                                     Inside Out
Name: Title, dtype: object
```

[86]:
```
#Plot profit vs budget
sns.set_style("whitegrid")
plt.figure(figsize=[8,6])
movies.plot.scatter(x='Profit', y='budget')
plt.title("Satter plot of Profit & Budget", fontsize= 15)
plt.show()
```

```
<Figure size 576x432 with 0 Axes>
```



The dataset contains the 100 best performing movies from the year 2010 to 2016. However scatter plot tells a different story. You can notice that there are some movies with negative profit. Although good movies do incur losses, but there appear to be quite a few movie with losses. What can be the reason behind this? Lets have a closer look at this by finding the movies with negative profit.

[87]:
```python
#Find the movies with negative profit

negative_profit = movies[movies["Profit"] < 0]
#negative_profit.sort_values(by='Profit', ascending=True)
negative_profit
```

[87]:
```
                       Title  title_year  budget       Gross  \
99      Tucker and Dale vs Evil        2010     5.0    0.223838
89                        Amour        2012     8.9    0.225377
56                         Rush        2013    38.0   26.903709
66                      Warrior        2011    25.0   13.651662
82                      Flipped        2010    14.0    1.752214
28           X-Men: First Class        2011   160.0  146.405371
46  Scott Pilgrim vs. the World        2010    60.0   31.494270
7                       Tangled        2010   260.0  200.807262
17            Edge of Tomorrow        2014   178.0  100.189501
39            The Little Prince        2015    81.2    1.339152
22                         Hugo        2011   170.0   73.820094

          actor_1_name        actor_2_name            actor_3_name  \
99      Katrina Bowden        Tyler Labine          Chelan Simmons
89    Isabelle Huppert     Emmanuelle Riva  Jean-Louis Trintignant
56     Chris Hemsworth        Olivia Wilde     Alexandra Maria Lara
66           Tom Hardy        Frank Grillo              Kevin Dunn
82    Madeline Carroll  Rebecca De Mornay             Aidan Quinn
28   Jennifer Lawrence  Michael Fassbender            Oliver Platt
46       Anna Kendrick       Kieran Culkin              Ellen Wong
7         Brad Garrett        Donna Murphy             M.C. Gainey
17          Tom Cruise         Lara Pulver             Noah Taylor
39        Jeff Bridges        James Franco          Mackenzie Foy
22  ChloÃ« Grace Moretz    Christopher Lee           Ray Winstone

    actor_1_facebook_likes  actor_2_facebook_likes  actor_3_facebook_likes  \
99                     948                   779.0                   440.0
89                     678                   432.0                   319.0
56                   26000                 10000.0                   471.0
66                   27000                   798.0                   581.0
82                    1000                   872.0                   767.0
28                   34000                 13000.0                  1000.0
46                   10000                  1000.0                   719.0
7                      799                   553.0                   284.0
17                   10000                   854.0                   509.0
39                   12000                 11000.0                  6000.0
```

22	17000	16000.0	1000.0

	IMDb_rating	genre_1	genre_2	genre_3	MetaCritic	Runtime	CVotes10 \
99	7.6	Comedy	Horror	NaN	65.0	124	16572
89	7.9	Drama	Romance	NaN	94.0	127	11093
56	8.1	Action	Biography	Drama	75.0	123	53667
66	8.2	Action	Drama	Sport	71.0	140	74983
82	7.7	Comedy	Drama	Romance	NaN	124	11354
28	7.8	Action	Adventure	Sci-Fi	65.0	132	64428
46	7.5	Action	Comedy	Romance	69.0	112	47292
7	7.8	Animation	Adventure	Comedy	71.0	124	56575
17	7.9	Action	Adventure	Sci-Fi	71.0	113	60383
39	7.8	Animation	Adventure	Drama	70.0	108	7565
22	7.5	Adventure	Drama	Family	83.0	126	29228

	CVotes09	CVotes08	CVotes07	CVotes06	CVotes05	CVotes04	CVotes03 \
99	19818	44460	35863	13456	4588	1684	855
89	15944	22942	14187	5945	2585	1188	710
56	90907	119603	57343	14948	4436	1625	803
66	96953	106673	52972	16668	5727	2353	1205
82	11050	20808	14372	5412	1848	664	321
28	96219	200144	129352	41945	12861	4799	2349
46	48976	79198	59689	28452	13451	6977	4254
7	54688	97207	70947	26805	8530	3043	1396
17	99596	175961	100724	28982	8145	2858	1368
39	7321	11668	8558	3370	1162	456	227
22	40728	77893	62936	27932	11179	4664	2674

	CVotes02	CVotes01	CVotesMale	CVotesFemale	CVotesU18	CVotesU18M \
99	479	848	106144	15113	219	198
89	534	995	49808	16719	121	95
56	633	1532	246354	35289	888	769
66	1050	2479	270734	31075	673	583
82	230	402	33714	22540	320	108
28	1448	3182	382107	80444	2075	1612
46	3069	6287	208417	45718	1022	791
7	805	1606	166088	97213	1950	1048
17	857	1661	341199	45790	1893	1584
39	158	293	22441	9552	226	128
22	1700	3023	179655	47795	884	616

	CVotesU18F	CVotes1829	CVotes1829M	CVotes1829F	CVotes3044	CVotes3044M \
99	20	52889	45169	7232	56379	49634
89	24	28593	20107	8167	28691	21990
56	115	138327	119048	17987	105712	92487
66	84	153824	136536	16000	117636	105144
82	207	32475	16916	15217	18576	12982

28	443	223309	176821	44428	185909	157332
46	221	129202	99363	28557	102986	88456
7	885	144744	81897	61390	89588	63534
17	292	172016	148814	21495	157476	138752
39	96	15959	10150	5610	12174	9280
22	250	105814	79064	25620	93126	76098

	CVotes3044F	CVotes45A	CVotes45AM	CVotes45AF	CVotes1000	CVotesUS	\
99	6156	8861	7645	1072	540	26213	
89	6269	7425	5803	1490	391	7959	
56	11717	20105	17096	2660	593	30738	
66	11019	15201	12960	1990	586	45342	
82	5338	3399	2616	721	275	7362	
28	26094	30217	25051	4691	780	87542	
46	13334	14195	12157	1778	677	56559	
7	24912	15318	11277	3805	622	47643	
17	16463	30965	26562	3820	691	55486	
39	2682	1899	1496	355	198	3678	
22	15694	22136	17667	4065	669	39127	

	CVotesnUS	VotesM	VotesF	VotesU18	VotesU18M	VotesU18F	Votes1829	\
99	73918	7.5	7.7	7.7	7.7	8.2	7.6	
89	46138	7.8	7.9	8.6	8.7	8.5	8.0	
56	168519	8.1	8.0	8.3	8.3	8.0	8.3	
66	176397	8.2	8.2	8.5	8.5	8.7	8.4	
82	36050	7.6	7.9	8.0	7.6	8.3	7.9	
28	257681	7.7	7.9	8.0	7.9	8.3	7.9	
46	150511	7.5	7.4	8.1	8.2	8.0	7.6	
7	148024	7.6	8.2	7.8	7.4	8.3	7.9	
17	217557	7.9	7.8	8.2	8.1	8.2	8.0	
39	19009	7.6	8.1	8.0	7.8	8.2	7.9	
22	142850	7.5	7.6	7.8	7.8	7.6	7.7	

	Votes1829M	Votes1829F	Votes3044	Votes3044M	Votes3044F	Votes45A	\
99	7.6	7.6	7.5	7.5	7.7	7.5	
89	8.0	7.9	7.7	7.7	7.9	7.9	
56	8.3	8.2	8.0	8.0	7.9	7.8	
66	8.4	8.4	8.0	8.0	8.0	7.7	
82	7.8	8.0	7.5	7.5	7.7	7.4	
28	7.9	8.0	7.7	7.6	7.8	7.6	
46	7.7	7.5	7.4	7.4	7.2	7.1	
7	7.7	8.2	7.6	7.5	8.0	7.7	
17	8.0	7.8	7.8	7.8	7.7	7.8	
39	7.8	8.2	7.6	7.5	7.9	7.5	
22	7.7	7.7	7.4	7.4	7.4	7.5	

	Votes45AM	Votes45AF	Votes1000	VotesUS	VotesnUS	content_rating	Country	\

99	7.4	7.7	7.1	7.7	7.5	R	Canada
89	7.8	8.1	7.2	7.9	7.8	PG-13	France
56	7.8	7.8	7.1	7.9	8.1	R	UK
66	7.7	7.5	7.1	8.2	8.1	PG-13	USA
82	7.3	7.6	6.4	7.5	7.7	PG	USA
28	7.5	7.7	7.3	7.8	7.7	PG-13	USA
46	7.1	7.0	6.6	7.8	7.4	PG-13	USA
7	7.6	7.9	6.9	7.9	7.7	PG	USA
17	7.8	7.8	7.5	8.0	7.8	PG-13	USA
39	7.4	7.9	6.6	7.7	7.7	PG	France
22	7.5	7.6	7.4	7.7	7.5	PG	USA

	Profit
99	-4.776162
89	-8.674623
56	-11.096291
66	-11.348338
82	-12.247786
28	-13.594629
46	-28.505730
7	-59.192738
17	-77.810499
39	-79.860848
22	-96.179906

`Checkpoint 1:` Can you spot the movie `Tangled` in the dataset? You may be aware of the movie 'Tangled'. Although its one of the highest grossing movies of all time, it has negative profit as per this result. If you cross check the gross values of this movie (link: https://www.imdb.com/title/tt0398286/), you can see that the gross in the dataset accounts only for the domestic gross and not the worldwide gross. This is true for may other movies also in the list.

-

1.2.3 Subtask 2.3: The General Audience and the Critics

You might have noticed the column `MetaCritic` in this dataset. This is a very popular website where an average score is determined through the scores given by the top-rated critics. Second, you also have another column `IMDb_rating` which tells you the IMDb rating of a movie. This rating is determined by taking the average of hundred-thousands of ratings from the general audience.

As a part of this subtask, you are required to find out the highest rated movies which have been liked by critics and audiences alike. 1. Firstly you will notice that the `MetaCritic` score is on a scale of `100` whereas the `IMDb_rating` is on a scale of 10. First convert the `MetaCritic` column to a scale of 10. 2. Now, to find out the movies which have been liked by both critics and audiences alike and also have a high rating overall, you need to - - Create a new column `Avg_rating` which will have the average of the `MetaCritic` and `Rating` columns - Retain only the movies in which the absolute difference(using abs() function) between the `IMDb_rating` and `Metacritic` columns is less than 0.5. Refer to this link to know how abs() funtion works - https://www.geeksforgeeks.org/abs-in-python/ . - Sort these values in a descending order of `Avg_rating` and retain only the movies with a rating equal to higher than 8 and store these movies in a new dataframe `UniversalAcclaim`.

[88]:
```python
# Change the scale of MetaCritic
movies['MetaCritic'] = movies['MetaCritic']/10
```

[89]:
```python
# Find the average ratings
movies['Avg_rating'] = movies[['MetaCritic', 'IMDb_rating']].mean(axis=1)
movies
```

[89]:
```
                                          Title  title_year  budget  \
97  Star Wars: Episode VII - The Force Awakens        2015   245.0
11                                 The Avengers        2012   220.0
47                                     Deadpool        2016    58.0
32              The Hunger Games: Catching Fire        2013   130.0
12                                  Toy Story 3        2010   200.0
..                                          …           …       …
46                    Scott Pilgrim vs. the World        2010    60.0
7                                       Tangled        2010   260.0
17                             Edge of Tomorrow        2014   178.0
39                             The Little Prince        2015    81.2
22                                         Hugo        2011   170.0

         Gross         actor_1_name        actor_2_name          actor_3_name  \
97  936.662225          Doug Walker          Rob Walker                     0
11  623.279547      Chris Hemsworth  Robert Downey Jr.     Scarlett Johansson
47  363.024263        Ryan Reynolds           Ed Skrein        Stefan Kapicic
32  424.645577    Jennifer Lawrence     Josh Hutcherson  Sandra Ellis Lafferty
12  414.984497            Tom Hanks  John Ratzenberger            Don Rickles
..         …                    …                  …                     …
46   31.494270        Anna Kendrick       Kieran Culkin            Ellen Wong
7   200.807262         Brad Garrett        Donna Murphy           M.C. Gainey
17  100.189501           Tom Cruise         Lara Pulver           Noah Taylor
39    1.339152         Jeff Bridges        James Franco        Mackenzie Foy
22   73.820094  ChloÃ« Grace Moretz    Christopher Lee          Ray Winstone

    actor_1_facebook_likes  actor_2_facebook_likes  actor_3_facebook_likes  \
97                     131                    12.0                     0.0
11                   26000                 21000.0                 19000.0
47                   16000                   805.0                   361.0
32                   34000                 14000.0                   523.0
12                   15000                  1000.0                   721.0
..                     …                      …                      …
46                   10000                  1000.0                   719.0
7                      799                   553.0                   284.0
17                   10000                   854.0                   509.0
39                   12000                 11000.0                  6000.0
```

22	17000	16000.0	1000.0

	IMDb_rating	genre_1	genre_2	genre_3	MetaCritic	Runtime	CVotes10	\
97	8.1	Action	Adventure	Fantasy	8.1	136	155391	
11	8.1	Action	Sci-Fi	NaN	6.9	143	260257	
47	8.0	Action	Adventure	Comedy	6.5	108	147467	
32	7.6	Action	Adventure	Mystery	7.6	146	85219	
12	8.3	Animation	Adventure	Comedy	9.2	103	139773	
..	…	…	…	…	…	…	…	
46	7.5	Action	Comedy	Romance	6.9	112	47292	
7	7.8	Animation	Adventure	Comedy	7.1	124	56575	
17	7.9	Action	Adventure	Sci-Fi	7.1	113	60383	
39	7.8	Animation	Adventure	Drama	7.0	108	7565	
22	7.5	Adventure	Drama	Family	8.3	126	29228	

	CVotes09	CVotes08	CVotes07	CVotes06	CVotes05	CVotes04	CVotes03	\
97	161810	166378	99402	40734	18060	8751	5970	
11	234203	264290	162604	67579	27957	12176	7201	
47	147966	170810	105717	41811	15510	7046	4273	
32	83874	150153	121748	50575	18571	7591	4094	
12	149992	158704	88289	31291	11850	4859	2932	
..	…	…	…	…	…	…	…	
46	48976	79198	59689	28452	13451	6977	4254	
7	54688	97207	70947	26805	8530	3043	1396	
17	99596	175961	100724	28982	8145	2858	1368	
39	7321	11668	8558	3370	1162	456	227	
22	40728	77893	62936	27932	11179	4664	2674	

	CVotes02	CVotes01	CVotesMale	CVotesFemale	CVotesU18	CVotesU18M	\
97	4489	15768	425971	68664	4722	3919	
11	4996	15528	691783	151617	4953	3767	
47	3037	8538	391955	79804	4598	3601	
32	2675	6978	307237	115421	3650	1956	
12	2119	6586	389014	98386	3202	2405	
..	…	…	…	…	…	…	
46	3069	6287	208417	45718	1022	791	
7	805	1606	166088	97213	1950	1048	
17	857	1661	341199	45790	1893	1584	
39	158	293	22441	9552	226	128	
22	1700	3023	179655	47795	884	616	

	CVotesU18F	CVotes1829	CVotes1829M	CVotes1829F	CVotes3044	CVotes3044M	\
97	768	220467	183671	34366	187138	162918	
11	1150	432999	343012	85465	295318	247617	
47	969	232840	186139	44316	159222	135428	
32	1664	218884	148652	67934	140683	109976	
12	776	260519	199962	58366	169886	140253	

```
..         …         …          …          …         …          …
46       221    129202      99363      28557    102986      88456
7        885    144744      81897      61390     89588      63534
17       292    172016     148814      21495    157476     138752
39        96     15959      10150       5610     12174       9280
22       250    105814      79064      25620     93126      76098

    CVotes3044F  CVotes45A  CVotes45AM  CVotes45AF  CVotes1000  CVotesUS  \
97        21362      42942       36441        5729         712     85141
11        43303      54282       44183        9138         842    145826
47        21521      28753       24218        4009         667     67933
32        28735      27789       21545        5771         693     68521
12        27658      32457       26171        5806         769    105490
..          …          …          …           …          …         …
46        13334      14195       12157        1778         677     56559
7         24912      15318       11277        3805         622     47643
17        16463      30965       26562        3820         691     55486
39         2682       1899        1496         355         198      3678
22        15694      22136       17667        4065         669     39127

    CVotesnUS  VotesM  VotesF  VotesU18  VotesU18M  VotesU18F  Votes1829  \
97     250769     8.0     8.3       8.5        8.5        8.6        8.2
11     423958     8.0     8.2       8.2        8.2        8.5        8.1
47     241138     8.0     8.1       8.4        8.4        8.6        8.1
32     221430     7.4     8.1       8.0        7.7        8.5        7.8
12     267692     8.3     8.3       8.2        8.3        8.0        8.4
..          …       …      …         …          …          …          …
46     150511     7.5     7.4       8.1        8.2        8.0        7.6
7      148024     7.6     8.2       7.8        7.4        8.3        7.9
17     217557     7.9     7.8       8.2        8.1        8.2        8.0
39      19009     7.6     8.1       8.0        7.8        8.2        7.9
22     142850     7.5     7.6       7.8        7.8        7.6        7.7

    Votes1829M  Votes1829F  Votes3044  Votes3044M  Votes3044F  Votes45A  \
97         8.2         8.3        8.0         7.9         8.2       7.9
11         8.1         8.3        8.0         8.0         8.1       7.9
47         8.1         8.1        7.9         7.9         7.9       7.8
32         7.6         8.2        7.3         7.2         7.9       7.3
12         8.5         8.4        8.2         8.2         8.3       8.1
..           …           …          …           …           …         …
46         7.7         7.5        7.4         7.4         7.2       7.1
7          7.7         8.2        7.6         7.5         8.0       7.7
17         8.0         7.8        7.8         7.8         7.7       7.8
39         7.8         8.2        7.6         7.5         7.9       7.5
22         7.7         7.7        7.4         7.4         7.4       7.5

    Votes45AM  Votes45AF  Votes1000  VotesUS  VotesnUS content_rating Country  \
```

```
97          7.8         8.2         7.7        8.2        7.9             PG-13      USA
11          7.9         8.1         7.4        8.3        7.9             PG-13      USA
47          7.8         7.9         7.3        8.1        7.9                 R      USA
32          7.2         7.9         6.7        7.7        7.4             PG-13      USA
12          8.1         8.1         8.1        8.5        8.3                 G      USA
..          …         …        …       …        …              …      …
46          7.1         7.0         6.6        7.8        7.4             PG-13      USA
7           7.6         7.9         6.9        7.9        7.7                PG      USA
17          7.8         7.8         7.5        8.0        7.8             PG-13      USA
39          7.4         7.9         6.6        7.7        7.7                PG   France
22          7.5         7.6         7.4        7.7        7.5                PG      USA

        Profit  Avg_rating
97  691.662225        8.10
11  403.279547        7.50
47  305.024263        7.25
32  294.645577        7.60
12  214.984497        8.75
..         …         …
46  -28.505730        7.20
7   -59.192738        7.45
17  -77.810499        7.50
39  -79.860848        7.40
22  -96.179906        7.90

[100 rows x 64 columns]
```

[90]:
```python
#Sort in descending order of average rating

movies.sort_values(by='Avg_rating', ascending = False , inplace = True)
```

[92]:
```python
movies
```

[92]:
```
                                          Title  title_year  budget  \
94                                      Boyhood        2014     4.0
69                             12 Years a Slave        2013    20.0
18                                   Inside Out        2015   175.0
0                                    La La Land        2016    30.0
4                         Manchester by the Sea        2016     9.0
..                                          …         …      …
28                         X-Men: First Class        2011   160.0
98  Harry Potter and the Deathly Hallows: Part I        2010   150.0
99                      Tucker and Dale vs Evil        2010     5.0
42                                         Fury        2014    68.0
44                               Les MisÃ©rables        2012    61.0

         Gross        actor_1_name        actor_2_name    actor_3_name  \
```

```
94   25.359200      Ellar Coltrane   Lorelei Linklater   Libby Villari
69   56.667870  QuvenzhanÃ© Wallis       Scoot McNairy    Taran Killam
18  356.454367         Amy Poehler        Mindy Kaling   Phyllis Smith
0   151.101803        Ryan Gosling          Emma Stone      Amiée Conn
4    47.695371       Casey Affleck  Michelle Williams    Kyle Chandler
..         …                   …                  …               …
28  146.405371   Jennifer Lawrence  Michael Fassbender    Oliver Platt
98  296.347721        Rupert Grint          Toby Jones    Alfred Enoch
99    0.223838      Katrina Bowden        Tyler Labine  Chelan Simmons
42   85.707116           Brad Pitt        Logan Lerman     Jim Parrack
44  148.775460        Hugh Jackman      Eddie Redmayne   Anne Hathaway

    actor_1_facebook_likes  actor_2_facebook_likes  actor_3_facebook_likes  \
94                     230                   193.0                   127.0
69                    2000                   660.0                   500.0
18                    1000                   767.0                   384.0
0                    14000                 19000.0                     NaN
4                      518                 71000.0                  3300.0
..                     …                       …                       …
28                   34000                 13000.0                  1000.0
98                   10000                  2000.0                  1000.0
99                     948                   779.0                   440.0
42                   11000                  8000.0                   697.0
44                   20000                 13000.0                 11000.0

    IMDb_rating    genre_1    genre_2  genre_3  MetaCritic  Runtime  CVotes10  \
94          7.9      Drama        NaN      NaN        10.0      165     49673
69          8.1  Biography      Drama  History         9.6      134     75556
18          8.2  Animation  Adventure   Comedy         9.4       95     87509
0           8.2     Comedy      Drama    Music         9.3      128     74245
4           7.9      Drama        NaN      NaN         9.6      137     18191
..          …          …          …        …          …        …        …
28          7.8     Action  Adventure   Sci-Fi         6.5      132     64428
98          7.7  Adventure     Family  Fantasy         6.5      146     68937
99          7.6     Comedy     Horror      NaN         6.5      124     16572
42          7.6     Action      Drama      War         6.4      134     36753
44          7.6      Drama    Musical  Romance         6.3      158     54268

    CVotes09  CVotes08  CVotes07  CVotes06  CVotes05  CVotes04  CVotes03  \
94     62055     76838     52238     23789     10431      4906      3071
69    126223    161460     83070     27231      9603      4021      2420
18    113244    119801     67153     24210      8542      3349      1872
0      71191     64640     38831     17377      8044      3998      2839
4      33532     46596     29626     11879      4539      1976      1233
..       …         …         …         …         …         …         …
28     96219    200144    129352     41945     12861      4799      2349
98     54947    102488     80465     31205     11792      4808      2454
```

99	19818	44460	35863	13456	4588	1684	855
42	54703	111271	82505	30231	10553	4303	2388
44	47750	63323	45160	22393	10744	5551	3484

	CVotes02	CVotes01	CVotesMale	CVotesFemale	CVotesU18	CVotesU18M	\
94	2248	5086	183807	51558	1393	995	
69	1785	4739	313823	82012	1837	1363	
18	1123	3450	244433	79081	3361	2294	
0	2407	6802	157693	56713	2675	1784	
4	888	1834	92452	22834	855	681	
..	…	…	…	…	…	…	
28	1448	3182	382107	80444	2075	1612	
98	1617	4522	223868	79506	1967	1310	
99	479	848	106144	15113	219	198	
42	1629	3246	238800	30746	1234	1028	
44	2490	5020	141014	73591	972	502	

	CVotesU18F	CVotes1829	CVotes1829M	CVotes1829F	CVotes3044	CVotes3044M	\
94	381	123006	92639	29076	81594	65261	
69	457	200910	153669	45301	138762	112943	
18	1040	170056	121574	46685	108560	86312	
0	868	113008	78998	32730	66058	50835	
4	166	55475	43467	11378	40645	32983	
..	…	…	…	…	…	…	
28	443	223309	176821	44428	185909	157332	
98	638	178871	126052	51112	94328	73103	
99	20	52889	45169	7232	56379	49634	
42	196	127986	110868	15886	99386	88611	
44	456	112609	68687	42720	69385	49760	

	CVotes3044F	CVotes45A	CVotes45AM	CVotes45AF	CVotes1000	CVotesUS	\
94	15118	17881	13995	3567	559	36433	
69	23895	29252	23072	5726	664	53328	
18	20516	18694	14910	3416	587	48297	
0	14165	15765	12148	3302	454	33360	
4	7053	11361	8862	2306	402	20287	
..	…	…	…	…	…	…	
28	26094	30217	25051	4691	780	87542	
98	20145	18077	14073	3750	734	56139	
99	6156	8861	7645	1072	540	26213	
42	9444	21524	18781	2392	548	35750	
44	18632	18404	12811	5282	623	42302	

	CVotesnUS	VotesM	VotesF	VotesU18	VotesU18M	VotesU18F	Votes1829	\
94	134679	8.0	7.7	8.1	8.1	8.0	8.1	
69	224519	8.1	8.1	8.4	8.4	8.5	8.2	
18	176446	8.2	8.2	8.4	8.4	8.3	8.3	

```
0      117987     8.2     8.1        8.9        9.0        8.7        8.4
4       65837     7.9     7.7        8.5        8.5        8.1        8.0
..          …       …       …          …          …          …          …
28     257681     7.7     7.9        8.0        7.9        8.3        7.9
98     180885     7.5     8.2        8.1        7.9        8.6        7.9
99      73918     7.5     7.7        7.7        7.7        8.2        7.6
42     148301     7.6     7.6        8.0        8.0        8.3        7.7
44     112787     7.5     7.9        7.9        7.6        8.4        7.8

    Votes1829M  Votes1829F  Votes3044  Votes3044M  Votes3044F  Votes45A  \
94         8.1         7.8        7.8         7.8         7.6       7.7
69         8.2         8.2        8.0         7.9         8.0       7.8
18         8.3         8.3        8.1         8.1         8.1       7.9
0          8.4         8.2        7.9         7.9         7.8       7.6
4          8.1         7.8        7.7         7.7         7.7       7.6
..           …           …          …           …           …         …
28         7.9         8.0        7.7         7.6         7.8       7.6
98         7.7         8.3        7.4         7.3         8.1       7.4
99         7.6         7.6        7.5         7.5         7.7       7.5
42         7.7         7.7        7.4         7.4         7.4       7.4
44         7.7         8.0        7.3         7.2         7.6       7.4

    Votes45AM  Votes45AF  Votes1000  VotesUS  VotesnUS content_rating Country  \
94        7.7        7.7        7.2      8.0       7.9              R     USA
69        7.8        8.1        7.7      8.3       8.0              R     USA
18        7.9        7.9        7.6      8.2       8.1             PG     USA
0         7.6        7.5        7.1      8.3       8.1          PG-13     USA
4         7.6        7.6        7.1      7.9       7.8              R     USA
..          …          …          …        …         …            …       …
28        7.5        7.7        7.3      7.8       7.7          PG-13     USA
98        7.3        8.0        6.7      7.9       7.5          PG-13      UK
99        7.4        7.7        7.1      7.7       7.5              R  Canada
42        7.4        7.4        6.8      7.6       7.5              R     USA
44        7.3        7.7        6.6      7.6       7.5          PG-13     USA

        Profit  Avg_rating
94   21.359200        8.95
69   36.667870        8.85
18  181.454367        8.80
0   121.101803        8.75
4    38.695371        8.75
..           …           …
28  -13.594629        7.15
98  146.347721        7.10
99   -4.776162        7.05
42   17.707116        7.00
44   87.775460        6.95
```

```
[100 rows x 64 columns]
```

[93]:
```python
# Find the movies with metacritic-rating < 0.5 and also with the average rating␣
↪of >8

UniversalAcclaim = movies[(abs(movies['MetaCritic'] - movies['IMDb_rating']) <␣
↪0.5) & (movies['Avg_rating']>=8)]
```

[94]:
```python
UniversalAcclaim
```

[94]:
```
                                         Title  title_year  budget  \
95                                    Whiplash        2014     3.3
35                            Django Unchained        2012   100.0
93                          Dallas Buyers Club        2013     5.0
97  Star Wars: Episode VII - The Force Awakens        2015   245.0
3                                      Arrival        2016    47.0
43                                   Gone Girl        2014    61.0
33                                 The Martian        2015   108.0

         Gross         actor_1_name     actor_2_name       actor_3_name  \
95   13.092000         J.K. Simmons  Melissa Benoist       Chris Mulkey
35  162.804648    Leonardo DiCaprio  Christoph Waltz       Ato Essandoh
93   27.296514  Matthew McConaughey  Jennifer Garner       Denis O'Hare
97  936.662225          Doug Walker       Rob Walker                  0
3   100.546139            Amy Adams    Jeremy Renner    Forest Whitaker
43  167.735396        Patrick Fugit        Sela Ward  Emily Ratajkowski
33  228.430993           Matt Damon    Donald Glover      Benedict Wong

    actor_1_facebook_likes  actor_2_facebook_likes  actor_3_facebook_likes  \
95                   24000                   970.0                   535.0
35                   29000                 11000.0                   265.0
93                   11000                  3000.0                   896.0
97                     131                    12.0                     0.0
3                    35000                  5300.0                     NaN
43                     835                   812.0                   625.0
33                   13000                   801.0                   372.0

    IMDb_rating    genre_1    genre_2  genre_3  MetaCritic  Runtime  CVotes10  \
95          8.5      Drama      Music      NaN         8.8      107    110404
35          8.4      Drama    Western      NaN         8.1      165    234824
93          8.0  Biography      Drama      NaN         8.4      117     37544
97          8.1     Action  Adventure  Fantasy         8.1      136    155391
3           8.0      Drama    Mystery   Sci-Fi         8.1      116     55533
43          8.1      Crime      Drama  Mystery         7.9      149     89539
33          8.0  Adventure      Drama   Sci-Fi         8.0      144     75560
```

	CVotes09	CVotes08	CVotes07	CVotes06	CVotes05	CVotes04	CVotes03	\
95	161864	132656	56007	16577	6031	2937	1859	
35	339329	286911	121445	38251	14227	6469	4149	
93	82276	145488	66156	16777	4582	1721	870	
97	161810	166378	99402	40734	18060	8751	5970	
3	87850	109536	65440	26913	10556	5057	3083	
43	177373	218018	103600	32989	11691	5285	3262	
33	139593	200315	102723	31179	9930	3815	2046	

	CVotes02	CVotes01	CVotesMale	CVotesFemale	CVotesU18	CVotesU18M	\
95	1263	2723	308900	71066	2878	2200	
35	3181	8065	695211	139226	3250	2726	
93	654	1588	231258	63266	864	650	
97	4489	15768	425971	68664	4722	3919	
3	2194	4734	237437	46272	1943	1544	
43	2247	5500	397571	113606	2286	1598	
33	1316	2907	359265	71421	3206	2543	

	CVotesU18F	CVotes1829	CVotes1829M	CVotes1829F	CVotes3044	CVotes3044M	\
95	660	205839	161853	41944	123712	102839	
35	501	410538	332690	74006	301231	253253	
93	205	145018	110493	32974	110555	88233	
97	768	220467	183671	34366	187138	162918	
3	376	126301	101741	23163	111985	95005	
43	665	260425	193602	64291	179552	144771	
33	638	200653	161765	36790	161073	136425	

	CVotes3044F	CVotes45A	CVotes45AM	CVotes45AF	CVotes1000	CVotesUS	\
95	19018	23345	19072	3812	590	49868	
35	43774	57463	47535	8962	816	123423	
93	20687	22093	16963	4743	598	42222	
97	21362	42942	36441	5729	712	85141	
3	15227	24027	20118	3440	537	42062	
43	32133	34696	27226	6840	689	70667	
33	22228	35406	29354	5409	671	61128	

	CVotesnUS	VotesM	VotesF	VotesU18	VotesU18M	VotesU18F	Votes1829	\
95	213952	8.5	8.4	9.0	9.1	8.9	8.6	
35	448126	8.4	8.4	8.8	8.9	8.5	8.6	
93	173002	7.9	8.1	8.2	8.2	8.1	8.1	
97	250769	8.0	8.3	8.5	8.5	8.6	8.2	
3	163774	7.9	8.0	8.6	8.6	8.4	8.2	
43	280587	8.1	8.1	8.5	8.6	8.4	8.3	
33	239125	8.0	8.1	8.4	8.4	8.5	8.1	

	Votes1829M	Votes1829F	Votes3044	Votes3044M	Votes3044F	Votes45A	\
95	8.7	8.5	8.3	8.3	8.2	8.1	

```
35         8.6        8.5        8.3        8.3        8.3      8.0
93         8.1        8.1        7.8        7.8        8.0      7.8
97         8.2        8.3        8.0        7.9        8.2      7.9
3          8.2        8.1        7.8        7.8        7.8      7.6
43         8.3        8.2        7.9        8.0        7.9      7.7
33         8.1        8.1        7.9        7.9        7.9      8.0

    Votes45AM  Votes45AF  Votes1000  VotesUS  VotesnUS content_rating Country  \
95        8.1        8.2        8.0      8.6       8.4              R     USA
35        8.0        8.1        7.8      8.4       8.4              R     USA
93        7.8        8.0        7.2      8.0       7.9              R     USA
97        7.8        8.2        7.7      8.2       7.9          PG-13     USA
3         7.6        7.7        7.3      8.0       7.9          PG-13     USA
43        7.7        7.7        7.6      8.1       8.1              R     USA
33        7.9        8.2        7.8      8.1       7.9          PG-13     USA

        Profit  Avg_rating
95    9.792000        8.65
35   62.804648        8.25
93   22.296514        8.20
97  691.662225        8.10
3    53.546139        8.05
43  106.735396        8.00
33  120.430993        8.00
```

Checkpoint 2: Can you spot a `Star Wars` movie in your final dataset?

-

1.2.4 Subtask 2.4: Find the Most Popular Trios - I

You're a producer looking to make a blockbuster movie. There will primarily be three lead roles in your movie and you wish to cast the most popular actors for it. Now, since you don't want to take a risk, you will cast a trio which has already acted in together in a movie before. The metric that you've chosen to check the popularity is the Facebook likes of each of these actors.

The dataframe has three columns to help you out for the same, viz. `actor_1_facebook_likes`, `actor_2_facebook_likes`, and `actor_3_facebook_likes`. Your objective is to find the trios which has the most number of Facebook likes combined. That is, the sum of `actor_1_facebook_likes`, `actor_2_facebook_likes` and `actor_3_facebook_likes` should be maximum. Find out the top 5 popular trios, and output their names in a list.

[95]:
```
# Write your code here
movies['Popularity'] = movies[['actor_1_facebook_likes',␣
 ↪'actor_2_facebook_likes','actor_3_facebook_likes']].sum(axis=1)
movies.sort_values(by='Popularity', ascending = False , inplace = True)
movies
```

[95]:

```
                                          Title  title_year  budget  \
2                                          Lion        2016    12.0
27                                    Inception        2010   160.0
14                     X-Men: Days of Future Past        2014   200.0
4                         Manchester by the Sea        2016     9.0
8                         The Dark Knight Rises        2012   250.0
..                                          …           …      …
94                                      Boyhood        2014     4.0
80                                  Ex Machina        2014    15.0
96                              Before Midnight        2013     3.0
97  Star Wars: Episode VII - The Force Awakens        2015   245.0
34                                      Gravity        2013   100.0

         Gross              actor_1_name        actor_2_name  \
2    51.738905                 Dev Patel       Nicole Kidman
27  292.568851         Leonardo DiCaprio           Tom Hardy
14  233.914986         Jennifer Lawrence      Peter Dinklage
4    47.695371             Casey Affleck  Michelle Williams
8   448.130642                 Tom Hardy      Christian Bale
..         …                        …                  …
94   25.359200            Ellar Coltrane   Lorelei Linklater
80   25.440971             Elina Alminas       Sonoya Mizuno
96    8.114507  Seamus Davey-Fitzpatrick        Ariane Labed
97  936.662225               Doug Walker          Rob Walker
34  274.084951            Phaldut Sharma       Basher Savage

              actor_3_name  actor_1_facebook_likes  actor_2_facebook_likes  \
2              Rooney Mara                   33000                 96000.0
27    Joseph Gordon-Levitt                   29000                 27000.0
14            Hugh Jackman                   34000                 22000.0
4            Kyle Chandler                     518                 71000.0
8     Joseph Gordon-Levitt                   27000                 23000.0
..                      …                      …                      …
94           Libby Villari                     230                   193.0
80           Corey Johnson                     149                   145.0
96  Athina Rachel Tsangari                     140                    63.0
97                       0                     131                    12.0
34              Amy Warren                      39                    23.0

    actor_3_facebook_likes  IMDb_rating    genre_1    genre_2   genre_3  \
2                   9800.0          8.1  Biography      Drama       NaN
27                 23000.0          8.8     Action  Adventure    Sci-Fi
14                 20000.0          8.0     Action  Adventure    Sci-Fi
4                   3300.0          7.9      Drama        NaN       NaN
8                  23000.0          8.4     Action   Thriller       NaN
..                      …          …         …         …       …
94                   127.0          7.9      Drama        NaN       NaN
```

80	123.0	7.7	Drama	Mystery	Sci-Fi
96	48.0	7.9	Drama	Romance	NaN
97	0.0	8.1	Action	Adventure	Fantasy
34	13.0	7.8	Drama	Sci-Fi	Thriller

	MetaCritic	Runtime	CVotes10	CVotes09	CVotes08	CVotes07	CVotes06	\
2	6.9	118	23325	29830	40564	20296	5842	
27	7.4	148	584839	485218	304457	130972	46393	
14	7.4	132	91765	127521	183578	104658	33027	
4	9.6	137	18191	33532	46596	29626	11879	
8	7.8	164	380589	341965	281426	134959	50406	
..	…	…	…	…	…	…	…	
94	10.0	165	49673	62055	76838	52238	23789	
80	7.8	108	29780	64769	123938	82736	28662	
96	9.4	109	16953	22109	31439	19251	8142	
97	8.1	136	155391	161810	166378	99402	40734	
34	9.6	91	89986	127616	169693	122275	57564	

	CVotes05	CVotes04	CVotes03	CVotes02	CVotes01	CVotesMale	\
2	1669	558	309	182	493	68921	
27	20595	10050	6631	5243	15365	1044318	
14	10059	3710	1903	1225	3301	370835	
4	4539	1976	1233	888	1834	92452	
8	20106	9589	5713	4073	11988	842343	
..	…	…	…	…	…	…	
94	10431	4906	3071	2248	5086	183807	
80	9579	3806	1951	1165	2182	237599	
96	3412	1649	1033	826	1745	67076	
97	18060	8751	5970	4489	15768	425971	
34	25393	12286	7868	5751	12473	427135	

	CVotesFemale	CVotesU18	CVotesU18M	CVotesU18F	CVotes1829	CVotes1829M	\
2	24977	702	477	220	42962	29729	
27	239796	5678	4462	1184	655187	512411	
14	71008	3038	2403	614	220178	179039	
4	22834	855	681	166	55475	43467	
8	143070	4726	4023	672	509635	425041	
..	…	…	…	…	…	…	
94	51558	1393	995	381	123006	92639	
80	41160	1154	899	242	126754	103143	
96	23823	208	138	66	43312	30016	
97	68664	4722	3919	768	220467	183671	
34	87618	2173	1684	468	233044	186837	

	CVotes1829F	CVotes3044	CVotes3044M	CVotes3044F	CVotes45A	CVotes45AM	\
2	12780	34297	26384	7413	9054	6714	
27	136770	472680	392845	73555	79634	65508	

14	39094	158607	135392	20927	26834	22460
4	11378	40645	32983	7053	11361	8862
8	79826	348324	299862	43434	55689	46968
..	…	…	…	…	…	…
94	29076	81594	65261	15118	17881	13995
80	22173	113021	97929	13354	23940	20773
96	12857	37072	28401	8189	7479	5891
97	34366	187138	162918	21362	42942	36441
34	43833	203844	171281	29467	44088	36531

	CVotes45AF	CVotes1000	CVotesUS	CVotesnUS	VotesM	VotesF	VotesU18	\
2	2184	298	13478	53931	8.0	8.4	8.3	
27	12795	885	212524	707266	8.8	8.7	9.1	
14	3884	710	67889	229049	8.0	8.1	8.4	
4	2306	402	20287	65837	7.9	7.7	8.5	
8	7741	840	160533	501687	8.5	8.4	8.6	
..	…	…	…	…	…	…	…	
94	3567	559	36433	134679	8.0	7.7	8.1	
80	2697	611	42556	164526	7.8	7.5	8.0	
96	1470	447	12382	59116	7.9	7.8	8.1	
97	5729	712	85141	250769	8.0	8.3	8.5	
34	6764	771	76797	292714	7.8	7.5	7.5	

	VotesU18M	VotesU18F	Votes1829	Votes1829M	Votes1829F	Votes3044	\
2	8.2	8.7	8.1	8.0	8.4	8.0	
27	9.1	9.0	9.0	9.0	8.8	8.7	
14	8.4	8.6	8.1	8.1	8.2	7.8	
4	8.5	8.1	8.0	8.1	7.8	7.7	
8	8.5	8.6	8.7	8.7	8.6	8.3	
..	…	…	…	…	…	…	
94	8.1	8.0	8.1	8.1	7.8	7.8	
80	8.1	7.8	7.9	7.9	7.6	7.6	
96	8.3	7.4	8.1	8.2	7.9	7.8	
97	8.5	8.6	8.2	8.2	8.3	8.0	
34	7.6	7.4	7.8	7.9	7.5	7.8	

	Votes3044M	Votes3044F	Votes45A	Votes45AM	Votes45AF	Votes1000	\
2	7.9	8.2	8.0	7.9	8.4	7.1	
27	8.7	8.5	8.1	8.1	8.0	8.2	
14	7.8	8.0	7.7	7.7	7.9	7.4	
4	7.7	7.7	7.6	7.6	7.6	7.1	
8	8.3	8.2	7.9	7.9	7.9	7.8	
..	…	…	…	…	…	…	
94	7.8	7.6	7.7	7.7	7.7	7.2	
80	7.6	7.4	7.6	7.6	7.4	7.5	
96	7.8	7.6	7.3	7.4	7.2	7.0	
97	7.9	8.2	7.9	7.8	8.2	7.7	

```
34          7.8          7.5        7.7         7.7         7.5         7.5

    VotesUS  VotesnUS  content_rating    Country      Profit  Avg_rating  \
2       8.1       8.0           PG-13  Australia   39.738905        7.50
27      8.7       8.8           PG-13        USA  132.568851        8.10
14      8.1       7.9           PG-13        USA   33.914986        7.70
4       7.9       7.8               R        USA   38.695371        8.75
8       8.4       8.4           PG-13        USA  198.130642        8.10
..      …         …             …        …         …           …
94      8.0       7.9               R        USA   21.359200        8.95
80      7.9       7.7               R         UK   10.440971        7.75
96      8.0       7.9               R        USA    5.114507        8.65
97      8.2       7.9           PG-13        USA  691.662225        8.10
34      7.9       7.8           PG-13         UK  174.084951        8.70

    Popularity
2     138800.0
27     79000.0
14     76000.0
4      74818.0
8      73000.0
..         …
94       550.0
80       417.0
96       251.0
97       143.0
34        75.0

[100 rows x 65 columns]
```

[96]:
```
top_5_popular_trios = movies.loc[:
 ↪8,['actor_1_name','actor_2_name','actor_3_name']].values.tolist()
top_5_popular_trios
```

[96]:
```
[['Dev Patel', 'Nicole Kidman', 'Rooney Mara'],
 ['Leonardo DiCaprio', 'Tom Hardy', 'Joseph Gordon-Levitt'],
 ['Jennifer Lawrence', 'Peter Dinklage', 'Hugh Jackman'],
 ['Casey Affleck', 'Michelle Williams ', 'Kyle Chandler'],
 ['Tom Hardy', 'Christian Bale', 'Joseph Gordon-Levitt']]
```

-

1.2.5 Subtask 2.5: Find the Most Popular Trios - II

In the previous subtask you found the popular trio based on the total number of facebook likes. Let's add a small condition to it and make sure that all three actors are popular. The condition is **none of the three actors' Facebook likes should be less than half of the other two**. For example, the following is a valid combo: - actor_1_facebook_likes: 70000 - actor_2_facebook_likes: 40000

- actor_3_facebook_likes: 50000

But the below one is not: - actor_1_facebook_likes: 70000 - actor_2_facebook_likes: 40000 - actor_3_facebook_likes: 30000

since in this case, `actor_3_facebook_likes` is 30000, which is less than half of `actor_1_facebook_likes`.

Having this condition ensures that you aren't getting any unpopular actor in your trio (since the total likes calculated in the previous question doesn't tell anything about the individual popularities of each actor in the trio.).

You can do a manual inspection of the top 5 popular trios you have found in the previous subtask and check how many of those trios satisfy this condition. Also, which is the most popular trio after applying the condition above?

Write your answers below.

- **`No. of trios that satisfy the above condition:`**
- **`Most popular trio after applying the condition:`**

`Optional:` Even though you are finding this out by a natural inspection of the dataframe, can you also achieve this through some *if-else* statements to incorporate this. You can try this out on your own time after you are done with the assignment.

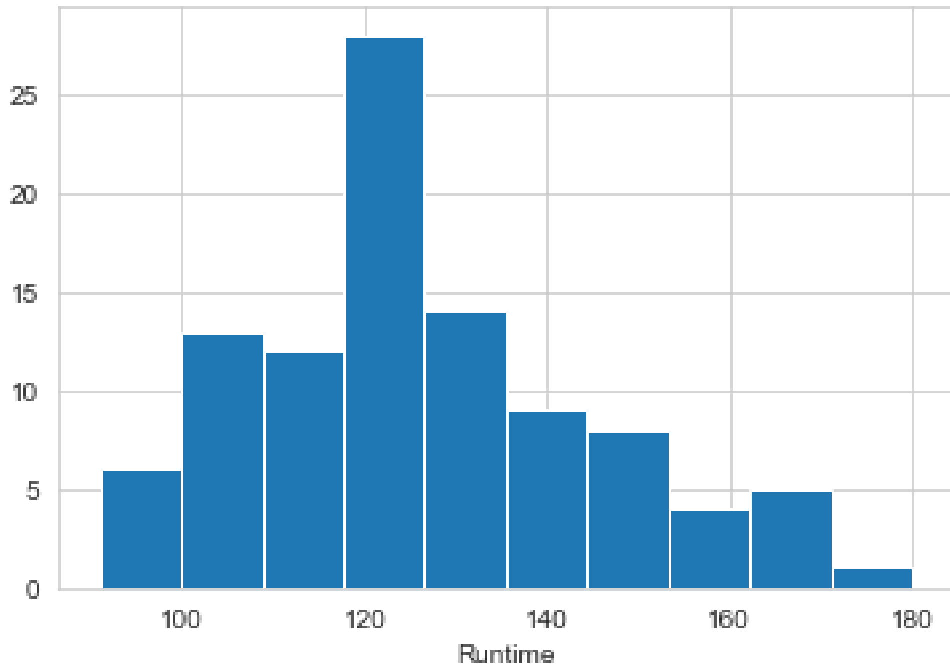
[97]:
```
# Your answer here (optional)
```

-

1.2.6 Subtask 2.6: Runtime Analysis

There is a column named `Runtime` in the dataframe which primarily shows the length of the movie. It might be intersting to see how this variable this distributed. Plot a `histogram` or `distplot` of seaborn to find the `Runtime` range most of the movies fall into.

[98]:
```
# Runtime histogram/density plot
plt.hist(movies.Runtime)
plt.xlabel("Runtime")
plt.show()
```



`Checkpoint 3:` Most of the movies appear to be sharply 2 hour-long.

-

1.2.7 Subtask 2.7: R-Rated Movies

Although R rated movies are restricted movies for the under 18 age group, still there are vote counts from that age group. Among all the R rated movies that have been voted by the under-18 age group, find the top 10 movies that have the highest number of votes i.e.`CVotesU18` from the `movies` dataframe. Store these in a dataframe named `PopularR`.

[99]:
```
# Write your code here
ContentR = movies[movies['content_rating'] == 'R']
PopularR = ContentR.sort_values(by='CVotesU18',ascending= False)
PopularR[0:10]
```

[99]:
```
                              Title  title_year  budget  \
47                         Deadpool        2016    58.0
36          The Wolf of Wall Street        2013   100.0
35                 Django Unchained        2012   100.0
29               Mad Max: Fury Road        2015   150.0
95                         Whiplash        2014     3.3
31                     The Revenant        2015   135.0
40                   Shutter Island        2010    80.0
43                        Gone Girl        2014    61.0
65         The Grand Budapest Hotel        2014    25.0
```

```
72  Birdman or (The Unexpected Virtue of Ignorance)        2014    18.0

         Gross       actor_1_name         actor_2_name       actor_3_name  \
47  363.024263      Ryan Reynolds            Ed Skrein     Stefan Kapicic
36  116.866727  Leonardo DiCaprio  Matthew McConaughey        Jon Favreau
35  162.804648  Leonardo DiCaprio      Christoph Waltz       Ato Essandoh
29  153.629485          Tom Hardy      Charlize Theron        ZoÃ« Kravitz
95   13.092000      J.K. Simmons      Melissa Benoist       Chris Mulkey
31  183.635922  Leonardo DiCaprio            Tom Hardy         Lukas Haas
40  127.968405  Leonardo DiCaprio        Joseph Sikora     Nellie Sciutto
43  167.735396      Patrick Fugit            Sela Ward  Emily Ratajkowski
65   59.073773        Bill Murray        Tom Wilkinson  F. Murray Abraham
72   42.335698         Emma Stone          Naomi Watts      Merritt Wever

    actor_1_facebook_likes  actor_2_facebook_likes  actor_3_facebook_likes  \
47                   16000                   805.0                   361.0
36                   29000                 11000.0                  4000.0
35                   29000                 11000.0                   265.0
29                   27000                  9000.0                   943.0
95                   24000                   970.0                   535.0
31                   29000                 27000.0                   733.0
40                   29000                   223.0                   163.0
43                     835                   812.0                   625.0
65                   13000                  1000.0                   670.0
72                   15000                  6000.0                   529.0

    IMDb_rating    genre_1    genre_2   genre_3  MetaCritic  Runtime  \
47          8.0     Action  Adventure    Comedy         6.5      108
36          8.2  Biography     Comedy     Crime         7.5      180
35          8.4      Drama    Western       NaN         8.1      165
29          8.1     Action  Adventure    Sci-Fi         9.0      120
95          8.5      Drama      Music       NaN         8.8      107
31          8.0  Adventure      Drama  Thriller         7.6      156
40          8.1    Mystery   Thriller       NaN         6.3      138
43          8.1      Crime      Drama   Mystery         7.9      149
65          8.1  Adventure     Comedy     Drama         8.8       99
72          7.8     Comedy      Drama       NaN         8.8      119

    CVotes10  CVotes09  CVotes08  CVotes07  CVotes06  CVotes05  CVotes04  \
47    147467    147966    170810    105717     41811     15510      7046
36    171660    236650    250667    129164     46715     18682      8674
35    234824    339329    286911    121445     38251     14227      6469
29    136194    158403    163494     97218     42636     19505      9932
95    110404    161864    132656     56007     16577      6031      2937
31     79977    121229    158019     91154     33492     12837      5571
40    150405    230844    278844    132349     45167     15615      7061
43     89539    177373    218018    103600     32989     11691      5285
```

65	84258	142011	168705	88086	31632	12023	5455
72	60209	94476	121637	80828	38373	19161	10116

	CVotes03	CVotes02	CVotes01	CVotesMale	CVotesFemale	CVotesU18	\
47	4273	3037	8538	391955	79804	4598	
36	5854	4258	9689	559564	123698	3622	
35	4149	3181	8065	695211	139226	3250	
29	6743	4930	10516	424435	69670	3159	
95	1859	1263	2723	308900	71066	2878	
31	3386	2320	4570	323938	61051	2619	
40	3780	2662	4703	570554	136360	2321	
43	3262	2247	5500	397571	113606	2286	
65	3196	2204	3971	332149	96997	2083	
72	6750	5378	11807	292808	63310	1891	

	CVotesU18M	CVotesU18F	CVotes1829	CVotes1829M	CVotes1829F	CVotes3044	\
47	3601	969	232840	186139	44316	159222	
36	2842	757	360841	286627	70874	227096	
35	2726	501	410538	332690	74006	301231	
29	2682	456	238202	198026	37751	183637	
95	2200	660	205839	161853	41944	123712	
31	2141	458	186003	152198	31926	138923	
40	1811	494	364786	283316	78332	263273	
43	1598	665	260425	193602	64291	179552	
65	1537	530	216106	158823	54913	153604	
72	1538	334	178850	142244	34666	129547	

	CVotes3044M	CVotes3044F	CVotes45A	CVotes45AM	CVotes45AF	CVotes1000	\
47	135428	21521	28753	24218	4009	667	
36	189110	34712	39996	32676	6629	730	
35	253253	43774	57463	47535	8962	816	
29	159520	21373	34848	29980	4209	726	
95	102839	19018	23345	19072	3812	590	
31	118140	18699	28582	23782	4269	624	
40	217923	42222	43235	35277	7256	840	
43	144771	32133	34696	27226	6840	689	
65	123255	27996	33156	26147	6375	679	
72	108049	19457	26016	21166	4329	656	

	CVotesUS	CVotesnUS	VotesM	VotesF	VotesU18	VotesU18M	VotesU18F	\
47	67933	241138	8.0	8.1	8.4	8.4	8.6	
36	89006	366829	8.2	7.8	8.6	8.7	8.0	
35	123423	448126	8.4	8.4	8.8	8.9	8.5	
29	73080	267084	8.1	7.9	8.5	8.6	8.4	
95	49868	213952	8.5	8.4	9.0	9.1	8.9	
31	51493	213741	8.0	7.9	8.5	8.5	8.2	
40	108244	419648	8.1	8.2	8.6	8.6	8.8	

```
43     70667     280587     8.1     8.1       8.5        8.6        8.4
65     58814     248067     8.1     8.2       8.6        8.6        8.5
72     52288     203731     7.8     7.5       8.5        8.6        7.9

    Votes1829  Votes1829M  Votes1829F  Votes3044  Votes3044M  Votes3044F  \
47        8.1         8.1         8.1        7.9         7.9         7.9
36        8.4         8.5         7.9        8.0         8.1         7.7
35        8.6         8.6         8.5        8.3         8.3         8.3
29        8.3         8.3         8.1        8.0         8.0         7.8
95        8.6         8.7         8.5        8.3         8.3         8.2
31        8.1         8.2         7.9        7.9         7.9         7.7
40        8.4         8.4         8.4        7.9         7.9         8.0
43        8.3         8.3         8.2        7.9         8.0         7.9
65        8.2         8.2         8.3        7.9         7.9         8.0
72        8.0         8.1         7.7        7.6         7.6         7.3

    Votes45A  Votes45AM  Votes45AF  Votes1000  VotesUS  VotesnUS  \
47       7.8        7.8        7.9        7.3      8.1       7.9
36       7.6        7.6        7.5        7.8      8.1       8.1
35       8.0        8.0        8.1        7.8      8.4       8.4
29       7.5        7.5        7.2        8.0      8.2       8.0
95       8.1        8.1        8.2        8.0      8.6       8.4
31       7.8        7.8        7.8        7.6      8.1       7.9
40       7.5        7.4        7.6        7.6      7.8       8.1
43       7.7        7.7        7.7        7.6      8.1       8.1
65       7.8        7.8        7.9        7.7      8.1       8.0
72       7.2        7.3        7.0        7.1      7.9       7.7

   content_rating    Country      Profit  Avg_rating  Popularity
47              R        USA  305.024263        7.25     17166.0
36              R        USA   16.866727        7.85     44000.0
35              R        USA   62.804648        8.25     40265.0
29              R  Australia    3.629485        8.55     36943.0
95              R        USA    9.792000        8.65     25505.0
31              R        USA   48.635922        7.80     56733.0
40              R        USA   47.968405        7.20     29386.0
43              R        USA  106.735396        8.00      2272.0
65              R        USA   34.073773        8.45     14670.0
72              R        USA   24.335698        8.30     21529.0
```

`Checkpoint 4:` Are these kids watching `Deadpool` a lot?

1.3 Task 3 : Demographic analysis

If you take a look at the last columns in the dataframe, most of these are related to demographics of the voters (in the last subtask, i.e., 2.8, you made use one of these columns - CVotesU18). We also have three genre columns indicating the genres of a particular movie. We will extensively use these columns for the third and the final stage of our assignment wherein we will analyse the voters across all demographics and also see how these vary across various genres. So without further ado, let's get started with `demographic analysis`.

-

1.3.1 Subtask 3.1 Combine the Dataframe by Genres

There are 3 columns in the dataframe - `genre_1`, `genre_2`, and `genre_3`. As a part of this subtask, you need to aggregate a few values over these 3 columns. 1. First create a new dataframe `df_by_genre` that contains `genre_1`, `genre_2`, and `genre_3` and all the columns related to **CVotes/Votes** from the `movies` data frame. There are 47 columns to be extracted in total. 2. Now, Add a column called `cnt` to the dataframe `df_by_genre` and initialize it to one. You will realise the use of this column by the end of this subtask. 3. First group the dataframe `df_by_genre` by `genre_1` and find the sum of all the numeric columns such as `cnt`, columns related to CVotes and Votes columns and store it in a dataframe `df_by_g1`. 4. Perform the same operation for `genre_2` and `genre_3` and store it dataframes `df_by_g2` and `df_by_g3` respectively. 5. Now that you have 3 dataframes performed by grouping over `genre_1`, `genre_2`, and `genre_3` separately, it's time to combine them. For this, add the three dataframes and store it in a new dataframe `df_add`, so that the corresponding values of Votes/CVotes get added for each genre.There is a function called `add()` in pandas which lets you do this. You can refer to this link to see how this function works. https://pandas.pydata.org/pandas-docs/version/0.23.4/generated/pandas.DataFrame.add.html 6. The column `cnt` on aggregation has basically kept the track of the number of occurences of each genre.Subset the genres that have atleast 10 movies into a new dataframe `genre_top10` based on the `cnt` column value. 7. Now, take the mean of all the numeric columns by dividing them with the column value `cnt` and store it back to the same dataframe. We will be using this dataframe for further analysis in this task unless it is explicitly mentioned to use the dataframe `movies`. 8. Since the number of votes can't be a fraction, type cast all the CVotes related columns to integers. Also, round off all the Votes related columns upto two digits after the decimal point.

[100]:
```python
# Create the dataframe df_by_genre

df_by_genre = pd.concat([movies.loc[:,'genre_1':'genre_3'],movies.loc[:
↪,'CVotes10':'VotesnUS']],axis=1)
df_by_genre.reset_index(drop= True , inplace= True)
```

[101]:
```python
df_by_genre.shape
```

[101]:
```
(100, 47)
```

[102]:
```python
# Create a column cnt and initialize it to 1

df_by_genre['cnt'] = 1
df_by_genre.head()
```

[102]:
```
     genre_1    genre_2 genre_3  CVotes10  CVotes09  CVotes08  CVotes07  \
0  Biography      Drama     NaN     23325     29830     40564     20296
1     Action  Adventure  Sci-Fi    584839    485218    304457    130972
2     Action  Adventure  Sci-Fi     91765    127521    183578    104658
```

```
3      Drama        NaN      NaN     18191     33532     46596     29626
4     Action   Thriller      NaN    380589    341965    281426    134959

   CVotes06  CVotes05  CVotes04  CVotes03  CVotes02  CVotes01  CVotesMale  \
0      5842      1669       558       309       182       493       68921
1     46393     20595     10050      6631      5243     15365     1044318
2     33027     10059      3710      1903      1225      3301      370835
3     11879      4539      1976      1233       888      1834       92452
4     50406     20106      9589      5713      4073     11988      842343

   CVotesFemale  CVotesU18  CVotesU18M  CVotesU18F  CVotes1829  CVotes1829M  \
0         24977        702         477         220       42962        29729
1        239796       5678        4462        1184      655187       512411
2         71008       3038        2403         614      220178       179039
3         22834        855         681         166       55475        43467
4        143070       4726        4023         672      509635       425041

   CVotes1829F  CVotes3044  CVotes3044M  CVotes3044F  CVotes45A  CVotes45AM  \
0        12780       34297        26384         7413       9054        6714
1       136770      472680       392845        73555      79634       65508
2        39094      158607       135392        20927      26834       22460
3        11378       40645        32983         7053      11361        8862
4        79826      348324       299862        43434      55689       46968

   CVotes45AF  CVotes1000  CVotesUS  CVotesnUS  VotesM  VotesF  VotesU18  \
0        2184         298     13478      53931     8.0     8.4       8.3
1       12795         885    212524     707266     8.8     8.7       9.1
2        3884         710     67889     229049     8.0     8.1       8.4
3        2306         402     20287      65837     7.9     7.7       8.5
4        7741         840    160533     501687     8.5     8.4       8.6

   VotesU18M  VotesU18F  Votes1829  Votes1829M  Votes1829F  Votes3044  \
0        8.2        8.7        8.1         8.0         8.4        8.0
1        9.1        9.0        9.0         9.0         8.8        8.7
2        8.4        8.6        8.1         8.1         8.2        7.8
3        8.5        8.1        8.0         8.1         7.8        7.7
4        8.5        8.6        8.7         8.7         8.6        8.3

   Votes3044M  Votes3044F  Votes45A  Votes45AM  Votes45AF  Votes1000  VotesUS  \
0         7.9         8.2       8.0        7.9        8.4        7.1      8.1
1         8.7         8.5       8.1        8.1        8.0        8.2      8.7
2         7.8         8.0       7.7        7.7        7.9        7.4      8.1
3         7.7         7.7       7.6        7.6        7.6        7.1      7.9
4         8.3         8.2       7.9        7.9        7.9        7.8      8.4

   VotesnUS  cnt
0       8.0    1
```

```
1        8.8    1
2        7.9    1
3        7.8    1
4        8.4    1
```

[103]:
```
# Group the movies by individual genres

df_by_g1 = df_by_genre.groupby('genre_1').sum()
df_by_g2 = df_by_genre.groupby('genre_2').sum()
df_by_g3 = df_by_genre.groupby('genre_3').sum()
```

[104]:
```
df_by_g1.head()
```

[104]:
```
           CVotes10  CVotes09  CVotes08  CVotes07  CVotes06  CVotes05  \
genre_1
Action      2928407   3261919   4247693   2662020    986774    364234
Adventure   1058779   1179818   1560541    966275    365486    136985
Animation    681562    798227   1153214    722782    251076     83069
Biography    666831   1088430   1654704    962977    306247    100005
Comedy       371217    496905    770395    518566    205434     81933

           CVotes04  CVotes03  CVotes02  CVotes01  CVotesMale  CVotesFemale  \
genre_1
Action       156150     89483     61975    162426     9994618       1858890
Adventure     58559     33174     22018     48100     3535070        802902
Animation     30718     15733     10026     25193     2282985        724844
Biography     38874     21536     15365     37469     3151311        790809
Comedy        35788     20965     15286     33241     1573480        527333

           CVotesU18  CVotesU18M  CVotesU18F  CVotes1829  CVotes1829M  \
genre_1
Action         67450       53291       13665     5736871      4679868
Adventure      23513       18128        5195     2153334      1685453
Animation      23835       16354        7307     1588630      1140646
Biography      17284       12914        4210     1953708      1505949
Comedy          8718        5778        2859     1050091       739846

           CVotes1829F  CVotes3044  CVotes3044M  CVotes3044F  CVotes45A  \
genre_1
Action         1001097     4449857      3815204       572731     831269
Adventure       446466     1585917      1323556       239846     325640
Animation       432464     1043387       834069       195268     179365
Biography       428103     1449651      1185900       242976     301036
Comedy          299041      787591       617610       158450     156738

           CVotes45AM  CVotes45AF  CVotes1000  CVotesUS  CVotesnUS  VotesM  \
genre_1
```

```
Action       696252     120596      18788  1964491   6411552  211.8
Adventure    266247      53656       8099   659263   2463964   94.1
Animation    142636      33699       6364   524406   1654665   86.4
Biography    237968      57899       7620   595106   2261929  102.3
Comedy       122213      31665       4960   344669   1232105   70.0

           VotesF  VotesU18  VotesU18M  VotesU18F  Votes1829  Votes1829M  \
genre_1
Action      213.2     221.3      220.8      223.5      215.7       215.6
Adventure    95.4      98.4       98.0       98.1       96.0        96.2
Animation    89.3      88.9       87.7       90.7       88.5        87.9
Biography   102.9     106.7      106.4      106.7      103.9       104.1
Comedy       70.2      73.3       73.1       73.4       71.5        71.6

           Votes1829F  Votes3044  Votes3044M  Votes3044F  Votes45A  Votes45AM  \
genre_1
Action          215.3      209.1       208.8       210.0     206.5      206.0
Adventure        96.4       92.7        92.6        93.5      92.0       91.6
Animation        90.2       85.4        84.9        87.8      84.5       84.1
Biography       103.4      100.8       100.7       101.3     100.5      100.0
Comedy           70.8       68.6        68.7        68.9      67.7       67.5

           Votes45AF  Votes1000  VotesUS  VotesnUS  cnt
genre_1
Action         209.0      197.2    215.8     209.5   27
Adventure       93.8       88.9     95.3      93.5   12
Animation       86.7       80.0     87.6      86.1   11
Biography      102.9       94.7    103.3     101.5   13
Comedy          68.7       62.7     70.9      69.4    9
```

[105]:

```python
# Add the grouped data frames and store it in a new data frame

df_add = df_by_g1.add(df_by_g2 , fill_value=0).add(df_by_g3 , fill_value=0)
```

[106]:

```python
# Extract genres with atleast 10 occurences

genre_top10 = df_add[df_add['cnt'] >= 10]
genre_top10.head(10)
```

[106]:

```
             CVotes10   CVotes09   CVotes08   CVotes07   CVotes06  CVotes05  \
Action      3166467.0  3547429.0  4677755.0  2922126.0  1075354.0  393484.0
Adventure   3594659.0  4014192.0  5262328.0  3281981.0  1212075.0  438970.0
Animation    681562.0   798227.0  1153214.0   722782.0   251076.0   83069.0
Biography    852003.0  1401608.0  2231078.0  1332980.0   425595.0  138648.0
Comedy      1383616.0  1774987.0  2506851.0  1591069.0   600287.0  226852.0
Crime        574526.0   967118.0  1419495.0   821390.0   278391.0   98690.0
Drama       3404438.0  4935375.0  7107053.0  4319700.0  1529356.0  552312.0
```

```
Romance      549959.0   689492.0  1069280.0   712841.0    281289.0  110901.0
Sci-Fi      2325284.0  2530855.0  3002994.0  1802098.0    671811.0  254175.0
Thriller    1081701.0  1465491.0  1993378.0  1175799.0    416046.0  149953.0

            CVotes04  CVotes03  CVotes02  CVotes01  CVotesMale  CVotesFemale  \
Action      166970.0   95004.0   65573.0  171247.0  10837034.0     2105410.0
Adventure   183070.0  103318.0   69737.0  173858.0  11759815.0     2705904.0
Animation    30718.0   15733.0   10026.0   25193.0   2282985.0      724844.0
Biography    53718.0   29510.0   20613.0   51297.0   4329471.0      986889.0
Comedy       97469.0   56218.0   39391.0   88367.0   5223033.0     1498288.0
Crime        42271.0   24713.0   16985.0   37217.0   2881292.0      586638.0
Drama       235475.0  135126.0   94185.0  211308.0  14668668.0     3572166.0
Romance      48913.0   27698.0   19200.0   40075.0   2114693.0      826446.0
Sci-Fi      111925.0   65904.0   46171.0  114435.0   7181436.0     1421515.0
Thriller     65281.0   37940.0   25767.0   57630.0   4339209.0      893747.0

            CVotesU18  CVotesU18M  CVotesU18F  CVotes1829  CVotes1829M  \
Action        76280.0     59424.0     16302.0   6314227.0    5105793.0
Adventure     95791.0     72215.0     22864.0   7165429.0    5578727.0
Animation     23835.0     16354.0      7307.0   1588630.0    1140646.0
Biography     20937.0     15962.0      4780.0   2603294.0    2052779.0
Comedy        37971.0     27099.0     10574.0   3404755.0    2507369.0
Crime         12866.0     10253.0      2503.0   1677551.0    1349593.0
Drama         80847.0     59481.0     20608.0   8871540.0    6838230.0
Romance       13276.0      6959.0      6168.0   1536614.0    1027323.0
Sci-Fi        51305.0     40494.0     10431.0   4161743.0    3351092.0
Thriller      21632.0     17252.0      4195.0   2541027.0    2035455.0

            CVotes1829F  CVotes3044  CVotes3044M  CVotes3044F  CVotes45A  \
Action        1146894.0   4824318.0    4117926.0     639437.0   894736.0
Adventure     1516049.0   5262641.0    4400217.0     788533.0  1012382.0
Animation      432464.0   1043387.0     834069.0     195268.0   179365.0
Biography      524830.0   1999316.0    1658844.0     312373.0   405909.0
Comedy         862715.0   2453650.0    1985974.0     433363.0   458960.0
Crime          311413.0   1338103.0    1129391.0     190218.0   251542.0
Drama         1943300.0   6897952.0    5696903.0    1102691.0  1404823.0
Romance        493050.0   1049990.0     809301.0     225561.0   204349.0
Sci-Fi         769577.0   3204214.0    2722405.0     435556.0   614167.0
Thriller       480861.0   1995184.0    1682474.0     284988.0   384675.0

            CVotes45AM  CVotes45AF  CVotes1000   CVotesUS   CVotesnUS  VotesM  \
Action        746866.0    132357.0     21123.0  2144895.0   7005964.0   243.2
Adventure     832607.0    162152.0     25157.0  2345901.0   7905326.0   298.4
Animation     142636.0     33699.0      6364.0   524406.0   1654665.0    86.4
Biography     324710.0     74280.0     10804.0   817016.0   3078088.0   141.1
Comedy        367532.0     83303.0     13934.0  1132079.0   3815002.0   180.2
Crime         206799.0     40242.0      6821.0   549216.0   1994718.0    86.6
```

```
Drama        1132442.0    247420.0     38086.0  2870687.0  10480222.0   509.8
Romance       156714.0     43959.0      7130.0   482839.0   1706303.0   100.7
Sci-Fi        512399.0     90784.0     12146.0  1371825.0   4625171.0   135.4
Thriller      317480.0     60502.0      8705.0   823702.0   2976280.0   102.3

           VotesF  VotesU18  VotesU18M  VotesU18F  Votes1829  Votes1829M  \
Action      245.6     253.8      252.8      256.8      247.8       247.5
Adventure   304.5     311.4      309.3      314.9      304.9       304.3
Animation    89.3      88.9       87.7       90.7       88.5        87.9
Biography   141.8     146.9      146.6      146.9      143.4       143.7
Comedy      181.2     187.4      186.9      186.9      183.6       183.8
Crime        84.9      90.6       90.9       87.7       87.9        88.2
Drama       510.3     533.6      532.2      529.7      519.0       519.7
Romance     101.8     105.8      104.8      106.1      102.9       103.0
Sci-Fi      135.3     140.9      141.0      141.6      137.8       137.9
Thriller    101.6     106.7      106.7      104.9      104.1       104.3

           Votes1829F  Votes3044  Votes3044M  Votes3044F  Votes45A  Votes45AM  \
Action          248.0      240.0       239.5       241.8     237.0      236.4
Adventure       307.3      294.6       293.7       299.2     291.7      290.4
Animation        90.2       85.4        84.9        87.8      84.5       84.1
Biography       142.8      139.1       138.9       139.8     138.5      137.9
Comedy          182.8      177.4       177.4       178.3     175.0      174.7
Crime            85.4       84.9        85.4        83.7      83.9       83.8
Drama           514.8      501.3       501.1       501.8     496.8      495.3
Romance         102.8       98.9        98.9        99.6      97.8       97.5
Sci-Fi          136.7      133.6       133.5       133.2     131.1      130.8
Thriller        102.4      100.6       100.7       100.1      99.6       99.3

           Votes45AF  Votes1000  VotesUS  VotesnUS   cnt
Action         240.4      226.2    247.6     240.6  31.0
Adventure      298.0      280.6    303.5     296.2  38.0
Animation       86.7       80.0     87.6      86.1  11.0
Biography      141.7      130.1    142.7     139.9  18.0
Comedy         177.1      165.4    182.6     178.9  23.0
Crime           84.5       81.3     87.8      85.8  11.0
Drama          503.2      469.5    515.9     506.0  65.0
Romance         98.9       89.9    101.8     100.1  13.0
Sci-Fi         131.5      127.9    137.5     134.0  17.0
Thriller       100.7       96.2    103.1     101.5  13.0
```

[107]:

```python
# Take the mean for every column by dividing with cnt

genre_top10.loc[:,'CVotes10':'VotesnUS'] = genre_top10.loc[:,'CVotes10':
 ↪'VotesnUS'].div(genre_top10['cnt'], axis= 0)
genre_top10.head()
```

[107]:

```
               CVotes10       CVotes09       CVotes08      CVotes07  \
Action     102144.096774  114433.193548  150895.322581  94262.129032
Adventure   94596.289474  105636.631579  138482.315789  86367.921053
Animation   61960.181818   72566.090909  104837.636364  65707.454545
Biography   47333.500000   77867.111111  123948.777778  74054.444444
Comedy      60157.217391   77173.347826  108993.521739  69176.913043

              CVotes06      CVotes05     CVotes04     CVotes03     CVotes02  \
Action     34688.838710  12693.032258  5386.129032  3064.645161  2115.258065
Adventure  31896.710526  11551.842105  4817.631579  2718.894737  1835.184211
Animation  22825.090909   7551.727273  2792.545455  1430.272727   911.454545
Biography  23644.166667   7702.666667  2984.333333  1639.444444  1145.166667
Comedy     26099.434783   9863.130435  4237.782609  2444.260870  1712.652174

             CVotes01    CVotesMale  CVotesFemale    CVotesU18   CVotesU18M  \
Action     5524.096774  349581.741935  67916.451613  2460.645161  1916.903226
Adventure  4575.210526  309468.815789  71208.000000  2520.815789  1900.394737
Animation  2290.272727  207544.090909  65894.909091  2166.818182  1486.727273
Biography  2849.833333  240526.166667  54827.166667  1163.166667   886.777778
Comedy     3842.043478  227088.391304  65142.956522  1650.913043  1178.217391

           CVotesU18F     CVotes1829    CVotes1829M   CVotes1829F  \
Action     525.870968  203684.741935  164703.000000  36996.580645
Adventure  601.684211  188563.921053  146808.605263  39896.026316
Animation  664.272727  144420.909091  103695.090909  39314.909091
Biography  265.555556  144627.444444  114043.277778  29157.222222
Comedy     459.739130  148032.826087  109016.043478  37509.347826

              CVotes3044    CVotes3044M   CVotes3044F     CVotes45A  \
Action     155623.161290  132836.322581  20627.000000  28862.451613
Adventure  138490.552632  115795.184211  20750.868421  26641.631579
Animation   94853.363636   75824.454545  17751.636364  16305.909091
Biography  111073.111111   92158.000000  17354.055556  22550.500000
Comedy     106680.434783   86346.695652  18841.869565  19954.782609

             CVotes45AM   CVotes45AF  CVotes1000      CVotesUS      CVotesnUS  \
Action     24092.451613  4269.580645  681.387097  69190.161290  225998.838710
Adventure  21910.710526  4267.157895  662.026316  61734.236842  208034.894737
Animation  12966.909091  3063.545455  578.545455  47673.272727  150424.090909
Biography  18039.444444  4126.666667  600.222222  45389.777778  171004.888889
Comedy     15979.652174  3621.869565  605.826087  49220.826087  165869.652174

             VotesM    VotesF  VotesU18  VotesU18M  VotesU18F  Votes1829  \
Action     7.845161  7.922581  8.187097   8.154839   8.283871   7.993548
Adventure  7.852632  8.013158  8.194737   8.139474   8.286842   8.023684
Animation  7.854545  8.118182  8.081818   7.972727   8.245455   8.045455
Biography  7.838889  7.877778  8.161111   8.144444   8.161111   7.966667
```

```
Comedy     7.834783  7.878261  8.147826   8.126087   8.126087   7.982609

           Votes1829M  Votes1829F  Votes3044  Votes3044M  Votes3044F  \
Action       7.983871    8.000000   7.741935    7.725806    7.800000
Adventure    8.007895    8.086842   7.752632    7.728947    7.873684
Animation    7.990909    8.200000   7.763636    7.718182    7.981818
Biography    7.983333    7.933333   7.727778    7.716667    7.766667
Comedy       7.991304    7.947826   7.713043    7.713043    7.752174

           Votes45A  Votes45AM  Votes45AF  Votes1000   VotesUS  VotesnUS   cnt
Action     7.645161   7.625806   7.754839   7.296774  7.987097  7.761290  31.0
Adventure  7.676316   7.642105   7.842105   7.384211  7.986842  7.794737  38.0
Animation  7.681818   7.645455   7.881818   7.272727  7.963636  7.827273  11.0
Biography  7.694444   7.661111   7.872222   7.227778  7.927778  7.772222  18.0
Comedy     7.608696   7.595652   7.700000   7.191304  7.939130  7.778261  23.0
```

[108]:
```
# Rounding off the columns of Votes to two decimals

genre_top10.loc[:,'VotesM':'VotesnUS'] = round(genre_top10.loc[:,'VotesM':
 ↪'VotesnUS'],ndigits=2)
genre_top10.head()
```

[108]:
```
                CVotes10       CVotes09       CVotes08      CVotes07  \
Action     102144.096774  114433.193548  150895.322581  94262.129032
Adventure   94596.289474  105636.631579  138482.315789  86367.921053
Animation   61960.181818   72566.090909  104837.636364  65707.454545
Biography   47333.500000   77867.111111  123948.777778  74054.444444
Comedy      60157.217391   77173.347826  108993.521739  69176.913043

               CVotes06      CVotes05     CVotes04     CVotes03     CVotes02  \
Action     34688.838710  12693.032258  5386.129032  3064.645161  2115.258065
Adventure  31896.710526  11551.842105  4817.631579  2718.894737  1835.184211
Animation  22825.090909   7551.727273  2792.545455  1430.272727   911.454545
Biography  23644.166667   7702.666667  2984.333333  1639.444444  1145.166667
Comedy     26099.434783   9863.130435  4237.782609  2444.260870  1712.652174

              CVotes01     CVotesMale  CVotesFemale    CVotesU18   CVotesU18M  \
Action     5524.096774  349581.741935  67916.451613  2460.645161  1916.903226
Adventure  4575.210526  309468.815789  71208.000000  2520.815789  1900.394737
Animation  2290.272727  207544.090909  65894.909091  2166.818182  1486.727273
Biography  2849.833333  240526.166667  54827.166667  1163.166667   886.777778
Comedy     3842.043478  227088.391304  65142.956522  1650.913043  1178.217391

           CVotesU18F     CVotes1829    CVotes1829M   CVotes1829F  \
Action     525.870968  203684.741935  164703.000000  36996.580645
Adventure  601.684211  188563.921053  146808.605263  39896.026316
Animation  664.272727  144420.909091  103695.090909  39314.909091
```

```
Biography   265.555556  144627.444444  114043.277778  29157.222222
Comedy      459.739130  148032.826087  109016.043478  37509.347826

              CVotes3044    CVotes3044M   CVotes3044F     CVotes45A  \
Action      155623.161290  132836.322581  20627.000000  28862.451613
Adventure   138490.552632  115795.184211  20750.868421  26641.631579
Animation    94853.363636   75824.454545  17751.636364  16305.909091
Biography   111073.111111   92158.000000  17354.055556  22550.500000
Comedy      106680.434783   86346.695652  18841.869565  19954.782609

             CVotes45AM   CVotes45AF  CVotes1000      CVotesUS      CVotesnUS  \
Action      24092.451613  4269.580645  681.387097  69190.161290  225998.838710
Adventure   21910.710526  4267.157895  662.026316  61734.236842  208034.894737
Animation   12966.909091  3063.545455  578.545455  47673.272727  150424.090909
Biography   18039.444444  4126.666667  600.222222  45389.777778  171004.888889
Comedy      15979.652174  3621.869565  605.826087  49220.826087  165869.652174

            VotesM  VotesF  VotesU18  VotesU18M  VotesU18F  Votes1829  \
Action        7.85    7.92      8.19       8.15       8.28       7.99
Adventure     7.85    8.01      8.19       8.14       8.29       8.02
Animation     7.85    8.12      8.08       7.97       8.25       8.05
Biography     7.84    7.88      8.16       8.14       8.16       7.97
Comedy        7.83    7.88      8.15       8.13       8.13       7.98

            Votes1829M  Votes1829F  Votes3044  Votes3044M  Votes3044F  \
Action            7.98        8.00       7.74        7.73        7.80
Adventure         8.01        8.09       7.75        7.73        7.87
Animation         7.99        8.20       7.76        7.72        7.98
Biography         7.98        7.93       7.73        7.72        7.77
Comedy            7.99        7.95       7.71        7.71        7.75

            Votes45A  Votes45AM  Votes45AF  Votes1000  VotesUS  VotesnUS   cnt
Action          7.65       7.63       7.75       7.30     7.99      7.76  31.0
Adventure       7.68       7.64       7.84       7.38     7.99      7.79  38.0
Animation       7.68       7.65       7.88       7.27     7.96      7.83  11.0
Biography       7.69       7.66       7.87       7.23     7.93      7.77  18.0
Comedy          7.61       7.60       7.70       7.19     7.94      7.78  23.0
```

[109]:

```python
# Converting CVotes to int type

for i in genre_top10.columns[0:27]:
    genre_top10[i] = genre_top10[i].astype(int)
genre_top10
```

[109]:

```
            CVotes10  CVotes09  CVotes08  CVotes07  CVotes06  CVotes05  \
Action        102144    114433    150895     94262     34688     12693
Adventure      94596    105636    138482     86367     31896     11551
```

Animation	61960	72566	104837	65707	22825	7551
Biography	47333	77867	123948	74054	23644	7702
Comedy	60157	77173	108993	69176	26099	9863
Crime	52229	87919	129045	74671	25308	8971
Drama	52375	75928	109339	66456	23528	8497
Romance	42304	53037	82252	54833	21637	8530
Sci-Fi	136781	148873	176646	106005	39518	14951
Thriller	83207	112730	153336	90446	32003	11534

	CVotes04	CVotes03	CVotes02	CVotes01	CVotesMale	CVotesFemale \
Action	5386	3064	2115	5524	349581	67916
Adventure	4817	2718	1835	4575	309468	71208
Animation	2792	1430	911	2290	207544	65894
Biography	2984	1639	1145	2849	240526	54827
Comedy	4237	2444	1712	3842	227088	65142
Crime	3842	2246	1544	3383	261935	53330
Drama	3622	2078	1449	3250	225671	54956
Romance	3762	2130	1476	3082	162668	63572
Sci-Fi	6583	3876	2715	6731	422437	83618
Thriller	5021	2918	1982	4433	333785	68749

	CVotesU18	CVotesU18M	CVotesU18F	CVotes1829	CVotes1829M \
Action	2460	1916	525	203684	164703
Adventure	2520	1900	601	188563	146808
Animation	2166	1486	664	144420	103695
Biography	1163	886	265	144627	114043
Comedy	1650	1178	459	148032	109016
Crime	1169	932	227	152504	122690
Drama	1243	915	317	136485	105203
Romance	1021	535	474	118201	79024
Sci-Fi	3017	2382	613	244808	197123
Thriller	1664	1327	322	195463	156573

	CVotes1829F	CVotes3044	CVotes3044M	CVotes3044F	CVotes45A \
Action	36996	155623	132836	20627	28862
Adventure	39896	138490	115795	20750	26641
Animation	39314	94853	75824	17751	16305
Biography	29157	111073	92158	17354	22550
Comedy	37509	106680	86346	18841	19954
Crime	28310	121645	102671	17292	22867
Drama	29896	106122	87644	16964	21612
Romance	37926	80768	62253	17350	15719
Sci-Fi	45269	188483	160141	25620	36127
Thriller	36989	153475	129421	21922	29590

	CVotes45AM	CVotes45AF	CVotes1000	CVotesUS	CVotesnUS	VotesM \
Action	24092	4269	681	69190	225998	7.85

```
Adventure       21910         4267          662     61734     208034    7.85
Animation       12966         3063          578     47673     150424    7.85
Biography       18039         4126          600     45389     171004    7.84
Comedy          15979         3621          605     49220     165869    7.83
Crime           18799         3658          620     49928     181338    7.87
Drama           17422         3806          585     44164     161234    7.84
Romance         12054         3381          548     37141     131254    7.75
Sci-Fi          30141         5340          714     80695     272068    7.96
Thriller        24421         4654          669     63361     228944    7.87

           VotesF  VotesU18  VotesU18M  VotesU18F  Votes1829  Votes1829M  \
Action       7.92      8.19       8.15       8.28       7.99        7.98
Adventure    8.01      8.19       8.14       8.29       8.02        8.01
Animation    8.12      8.08       7.97       8.25       8.05        7.99
Biography    7.88      8.16       8.14       8.16       7.97        7.98
Comedy       7.88      8.15       8.13       8.13       7.98        7.99
Crime        7.72      8.24       8.26       7.97       7.99        8.02
Drama        7.85      8.21       8.19       8.15       7.98        8.00
Romance      7.83      8.14       8.06       8.16       7.92        7.92
Sci-Fi       7.96      8.29       8.29       8.33       8.11        8.11
Thriller     7.82      8.21       8.21       8.07       8.01        8.02

           Votes1829F  Votes3044  Votes3044M  Votes3044F  Votes45A  Votes45AM  \
Action           8.00       7.74        7.73        7.80      7.65       7.63
Adventure        8.09       7.75        7.73        7.87      7.68       7.64
Animation        8.20       7.76        7.72        7.98      7.68       7.65
Biography        7.93       7.73        7.72        7.77      7.69       7.66
Comedy           7.95       7.71        7.71        7.75      7.61       7.60
Crime            7.76       7.72        7.76        7.61      7.63       7.62
Drama            7.92       7.71        7.71        7.72      7.64       7.62
Romance          7.91       7.61        7.61        7.66      7.52       7.50
Sci-Fi           8.04       7.86        7.85        7.84      7.71       7.69
Thriller         7.88       7.74        7.75        7.70      7.66       7.64

           Votes45AF  Votes1000  VotesUS  VotesnUS   cnt
Action          7.75       7.30     7.99      7.76  31.0
Adventure       7.84       7.38     7.99      7.79  38.0
Animation       7.88       7.27     7.96      7.83  11.0
Biography       7.87       7.23     7.93      7.77  18.0
Comedy          7.70       7.19     7.94      7.78  23.0
Crime           7.68       7.39     7.98      7.80  11.0
Drama           7.74       7.22     7.94      7.78  65.0
Romance         7.61       6.92     7.83      7.70  13.0
Sci-Fi          7.74       7.52     8.09      7.88  17.0
Thriller        7.75       7.40     7.93      7.81  13.0
```

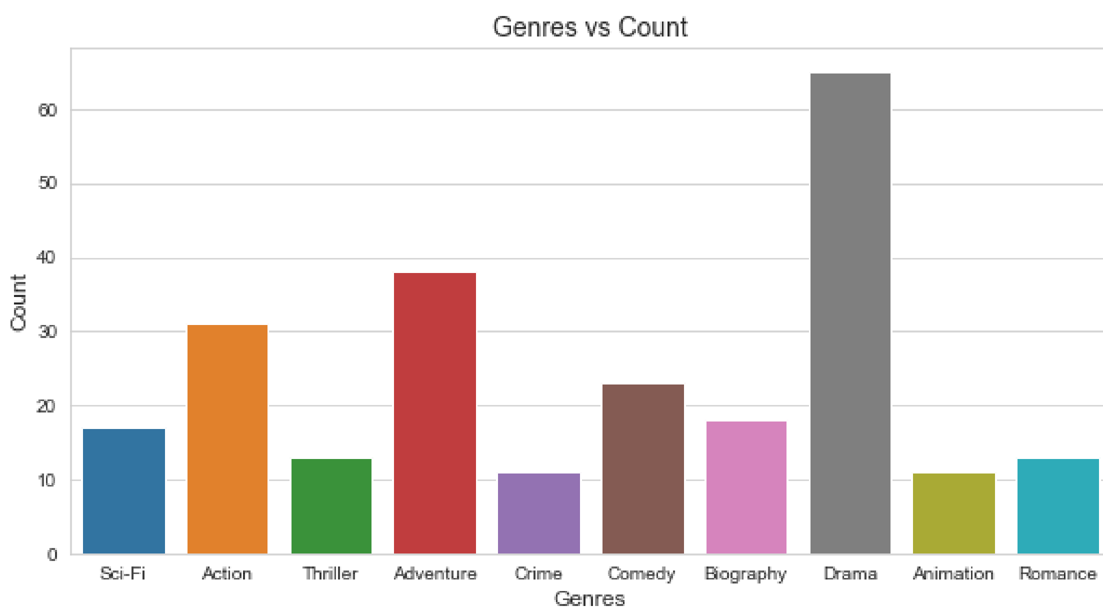
If you take a look at the final dataframe that you have gotten, you will see that you now have the complete information about all the demographic (Votes- and CVotes-related) columns across the top 10 genres. We can use this dataset to extract exciting insights about the voters!

-

1.3.2 Subtask 3.2: Genre Counts!

Now let's derive some insights from this data frame. Make a bar chart plotting different genres vs cnt using seaborn.

[129]:
```
# Countplot for genres
plt.figure(figsize=(10,5))
sns.barplot(x=genre_top10.index.values, y=genre_top10['cnt'])
plt.xlabel("Genres", fontsize= 12)
plt.ylabel("Count", fontsize= 12)
plt.title("Genres vs Count", fontsize= 14)
plt.show()
```



`Checkpoint 5:` Is the bar for `Drama` the tallest?

-

1.3.3 Subtask 3.3: Gender and Genre

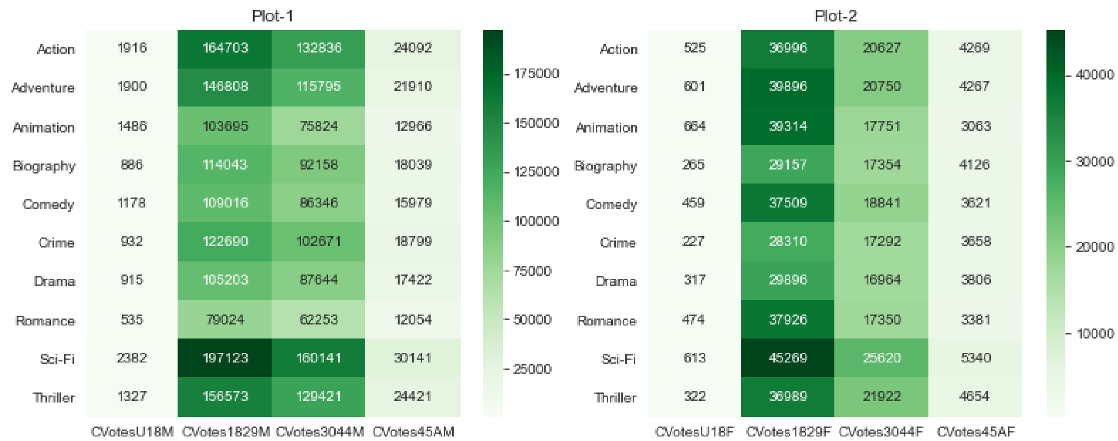
If you have closely looked at the Votes- and CVotes-related columns, you might have noticed the suffixes `F` and `M` indicating Female and Male. Since we have the vote counts for both males and females, across various age groups, let's now see how the popularity of genres vary between the two genders in the dataframe.

1. Make the first heatmap to see how the average number of votes of males is varying across the genres. Use seaborn heatmap for this analysis. The X-axis should contain the four age-groups for males, i.e., `CVotesU18M`,`CVotes1829M`, `CVotes3044M`, and `CVotes45AM`. The Y-axis will have the genres and the annotation in the heatmap tell the average number of votes for that age-male group.
2. Make the second heatmap to see how the average number of votes of females is varying across the genres. Use seaborn heatmap for this analysis. The X-axis should contain the four age-groups for females, i.e., `CVotesU18F`,`CVotes1829F`, `CVotes3044F`, and `CVotes45AF`. The Y-axis will have the genres and the annotation in the heatmap tell the average number of votes for that age-female group.
3. Make sure that you plot these heatmaps side by side using `subplots` so that you can easily compare the two genders and derive insights.
4. Write your any three inferences from this plot. You can make use of the previous bar plot also here for better insights. Refer to this link-https://seaborn.pydata.org/generated/seaborn.heatmap.html. You might have to plot something similar to the fifth chart in this page (You have to plot two such heatmaps side by side).
5. Repeat subtasks 1 to 4, but now instead of taking the CVotes-related columns, you need to do the same process for the Votes-related columns. These heatmaps will show you how the two genders have rated movies across various genres.

You might need the below link for formatting your heatmap. https://stackoverflow.com/questions/56942670/matplotlib-seaborn-first-and-last-row-cut-in-half-of-heatmap-plot

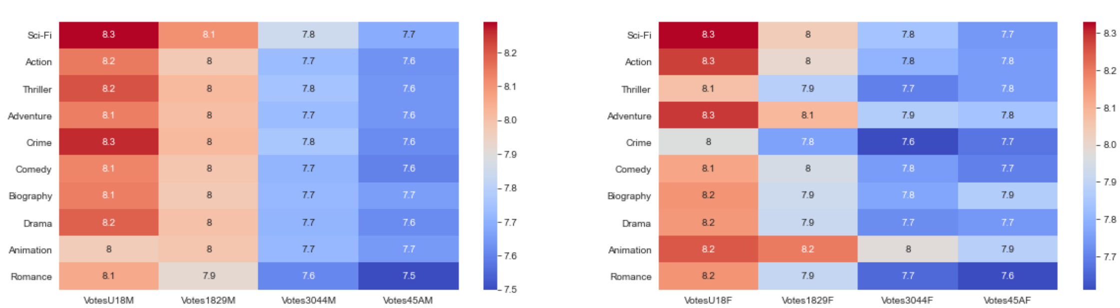
- Note : Use `genre_top10` dataframe for this subtask

[111]:
```
# 1st set of heat maps for CVotes-related columns
plt.figure(figsize=(13,5))
plt.subplot(1,2,1)
sns.
 ↪heatmap(genre_top10[['CVotesU18M','CVotes1829M','CVotes3044M','CVotes45AM']],cmap=␣
 ↪'Greens', annot= True, fmt= 'd')
plt.title('Plot-1')

plt.subplot(1,2,2)
sns.
 ↪heatmap(genre_top10[['CVotesU18F','CVotes1829F','CVotes3044F','CVotes45AF']],cmap=␣
 ↪'Greens', annot= True, fmt= 'd')
plt.title('Plot-2')
plt.show()
```



Inferences: A few inferences that can be seen from the heatmap above is that males have voted more than females, and Sci-Fi appears to be most popular among the 18-29 age group irrespective of their gender. What more can you infer from the two heatmaps that you have plotted? Write your three inferences/observations below: - Inference 1: Romance is the least voted by Male irrespective of their age compared to other genres(only male gender is considered). - Inference 2: Age ranging from 18–44 have voted the most for both Males and Females. - Inference 3: Sci-Fi has the highest number of votes Even if it has a lesser count compared to other genres (previous bar chart). - Inference 4: Movies seem to be watched less by U18 and 45A irrespective of their age or they watched but didn't vote

[127]:
```python
# 2ndset of heat maps for Votes-related columns

plt.figure(figsize=(20,5))

VoteM = genre_top10[['VotesU18M','Votes1829M','Votes3044M','Votes45AM']]
VoteF = genre_top10[['VotesU18F','Votes1829F','Votes3044F','Votes45AF']]

plt.subplot(1,2,1)
sns.heatmap(VoteM ,annot= True,cmap= 'coolwarm')
plt.subplot(1,2,2)
sns.heatmap(VoteF, annot= True,cmap= 'coolwarm')

plt.show()
```



Inferences: Sci-Fi appears to be the highest rated genre in the age group of U18 for both males and females. Also, females in this age group have rated it a bit higher than the males in the same age group. What more can you infer from the two heatmaps that you have plotted? Write your three inferences/observations below: - Inference 1: From the above Heatmap it is observed that females of all age likes animation movies - Inference 2: in general,Romance genre are watched less or voted less by males but the movies are good as they are rated well - irrespective of gender especially for U18 - Inference 3: Age ranging from 30–45 have their average rating to different genres is around 7.7 to 7.8(For both Male & Female).It seems to be that as age increses, you do it more professionaly(critic).

-

1.3.4 Subtask 3.4: US vs non-US Cross Analysis

The dataset contains both the US and non-US movies. Let's analyse how both the US and the non-US voters have responded to the US and the non-US movies.

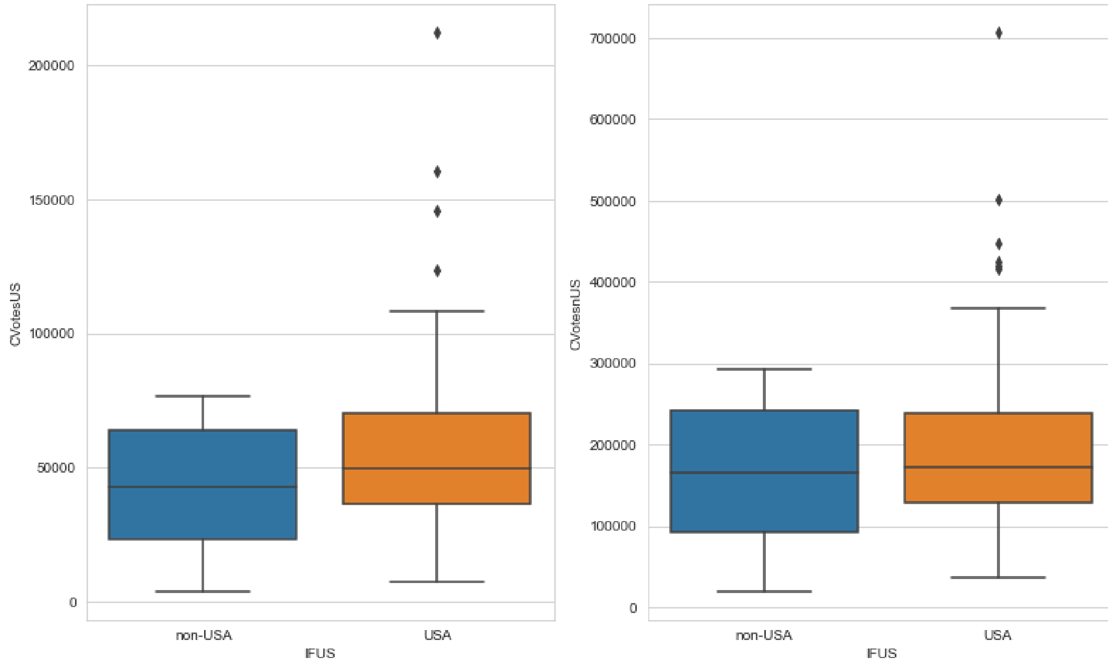
1. Create a column `IFUS` in the dataframe `movies`. The column `IFUS` should contain the value "USA" if the `Country` of the movie is "USA". For all other countries other than the USA, `IFUS` should contain the value `non-USA`.
2. Now make a boxplot that shows how the number of votes from the US people i.e. `CVotesUS` is varying for the US and non-US movies. Make use of the column `IFUS` to make this plot. Similarly, make another subplot that shows how non US voters have voted for the US and non-US movies by plotting `CVotesnUS` for both the US and non-US movies. Write any of your two inferences/observations from these plots.
3. Again do a similar analysis but with the ratings. Make a boxplot that shows how the ratings from the US people i.e. `VotesUS` is varying for the US and non-US movies. Similarly, make another subplot that shows how `VotesnUS` is varying for the US and non-US movies. Write any of your two inferences/observations from these plots.

Note : Use `movies` dataframe for this subtask. Make use of this documention to format your boxplot - https://seaborn.pydata.org/generated/seaborn.boxplot.html

[113]:
```
# Creating IFUS column

movies['IFUS'] = movies['Country'].apply(lambda x:'USA' if x == 'USA' else␣
↪'non-USA')
```

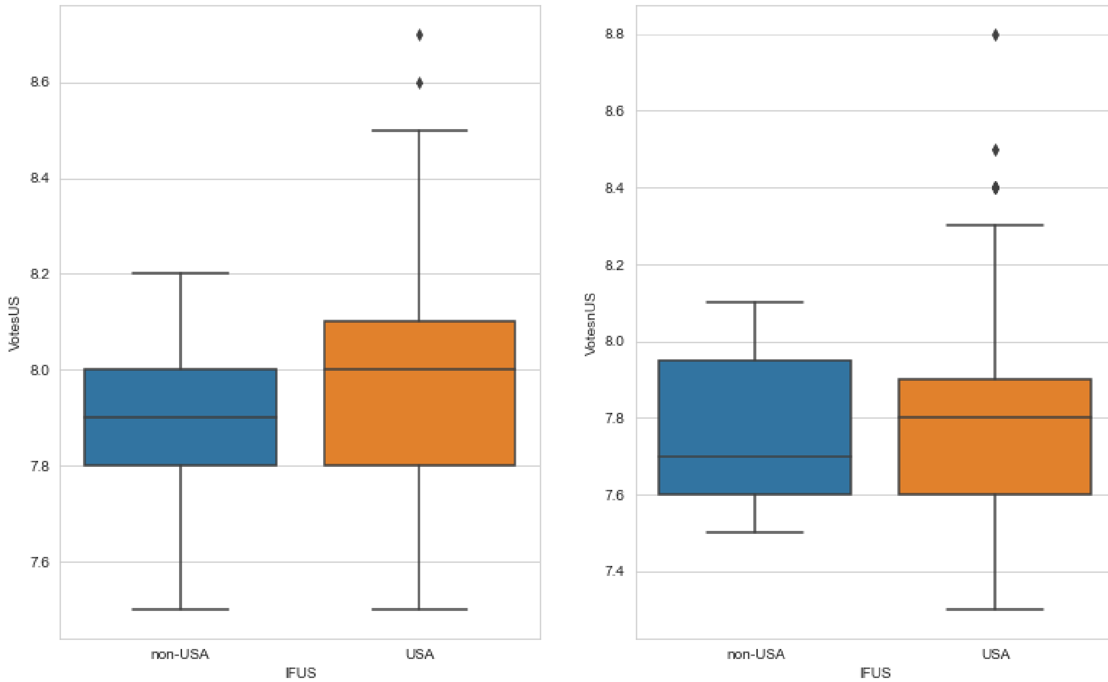
[114]:
```
# Box plot - 1: CVotesUS(y) vs IFUS(x)

plt.figure(figsize=(13,8))
plt.subplot(1,2,1)
sns.boxplot(x='IFUS' , y='CVotesUS',data = movies)

plt.subplot(1,2,2)
sns.boxplot(x='IFUS' , y='CVotesnUS',data = movies)
```

```
plt.show()
```



Inferences: Write your two inferences/observations below: - Inference 1: We can see some outliers for USA movies in both the plots. - Inference 2: irrespective of the origin, we can observe the number of votes on average is more from non-USA people as compared to USA people.

[125]:
```
# Box plot - 2: VotesUS(y) vs IFUS(x)

plt.figure(figsize=(13,8))
plt.subplot(1,2,1)
sns.boxplot(x='IFUS' , y='VotesUS',data = movies)

plt.subplot(1,2,2)
sns.boxplot(x='IFUS' , y='VotesnUS',data = movies)
plt.show()
```



Inferences: Write your two inferences/observations below: - Inference 1: In both the plots, rating of non-USA movies significantly is less than that of USA movies. - Inference 2: We can see the median rating is higher (around 7.9 to 8) from USA people compared to that from non- USA people.

-

1.3.5 Subtask 3.5: Top 1000 Voters Vs Genres

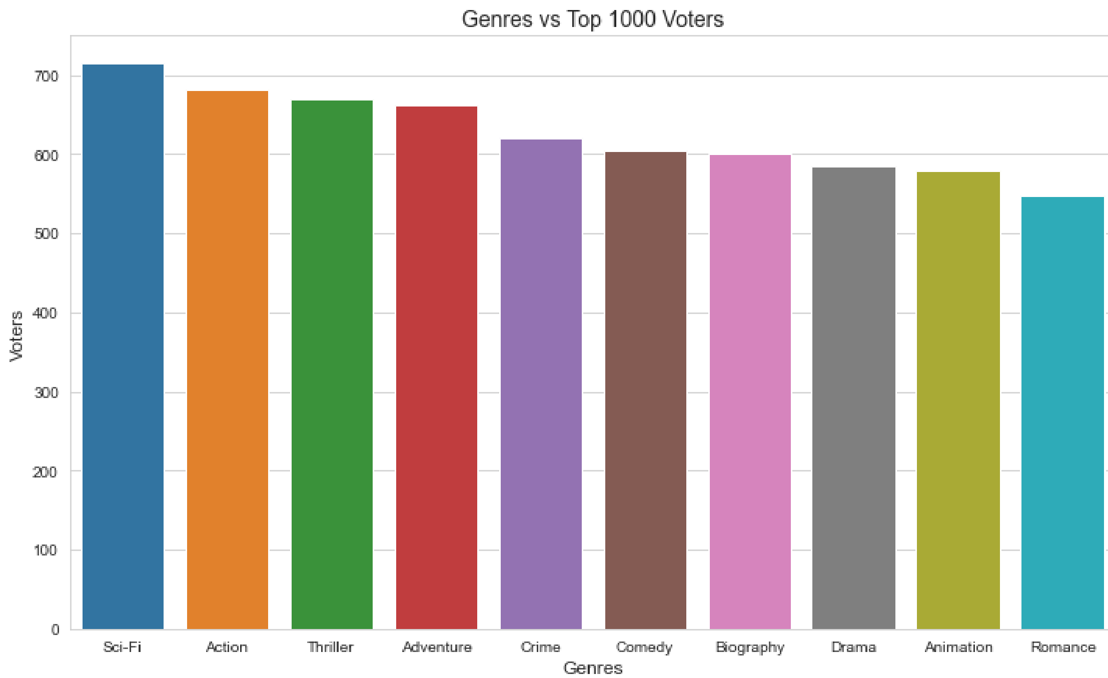
You might have also observed the column `CVotes1000`. This column represents the top 1000 voters on IMDb and gives the count for the number of these voters who have voted for a particular movie. Let's see how these top 1000 voters have voted across the genres.

1. Sort the dataframe genre_top10 based on the value of `CVotes1000`in a descending order.
2. Make a seaborn barplot for `genre` vs `CVotes1000`.
3. Write your inferences. You can also try to relate it with the heatmaps you did in the previous subtasks.

[116]:
```
# Sorting by CVotes1000

genre_top10.sort_values(by='CVotes1000', ascending = False, inplace= True)
```

[123]:
```
# Bar plot

plt.figure(figsize=(12,7))
sns.barplot(x=genre_top10.index.values, y=genre_top10['CVotes1000'])
plt.xlabel("Genres", fontsize= 12)
```

```
plt.ylabel("Voters", fontsize= 12)
plt.title("Genres vs Top 1000 Voters", fontsize= 14)
plt.show()
```



`Inferences:` The Romance genre has been voted the least and Sci-Fi is the most popular amongst the top 1000 voters.We can see that even if Sci-Fi has very few movies in the data set (as seen in the bar chart of count) still they got the highest rating across genders and the most number of votes from top 1000 IMDB Voters.

`Checkpoint 6:` The genre `Romance` seems to be most unpopular among the top 1000 voters.

With the above subtask, your assignment is over. In your free time, do explore the dataset further on your own and see what kind of other insights you can get across various other columns.