untitled26

June 29, 2023

[1]:
```
import pandas as pd
import numpy as np
```

[2]:
```
import warnings
warnings.filterwarnings('ignore')
```

1 Task 1: Reading and Inspection

Subtask 1.1: Import and read Import and read the movie database. Store it in a variable called movies

[4]:
```
path='C:\\Users\\surwa\\Downloads\\'
movies= pd.read_csv(path+ 'Movie+Assignment+Data (1).csv')
```

[5]:
```
movies.head()
```

[5]:

	color	director_name	num_critic_for_reviews	duration	\
0	Color	James Cameron	723.0	178.0	
1	Color	Gore Verbinski	302.0	169.0	
2	Color	Sam Mendes	602.0	148.0	
3	Color	Christopher Nolan	813.0	164.0	
4	NaN	Doug Walker	NaN	NaN	

	director_facebook_likes	actor_3_facebook_likes	actor_2_name	\
0	0.0	855.0	Joel David Moore	
1	563.0	1000.0	Orlando Bloom	
2	0.0	161.0	Rory Kinnear	
3	22000.0	23000.0	Christian Bale	
4	131.0	NaN	Rob Walker	

	actor_1_facebook_likes	gross	genres	…	\
0	1000.0	760505847.0	Action|Adventure|Fantasy|Sci-Fi	…	
1	40000.0	309404152.0	Action|Adventure|Fantasy	…	
2	11000.0	200074175.0	Action|Adventure|Thriller	…	
3	27000.0	448130642.0	Action|Thriller	…	
4	131.0	NaN	Documentary	…	

```
  num_user_for_reviews language  country  content_rating       budget  \
0               3054.0  English      USA           PG-13  237000000.0
1               1238.0  English      USA           PG-13  300000000.0
2                994.0  English       UK           PG-13  245000000.0
3               2701.0  English      USA           PG-13  250000000.0
4                  NaN      NaN      NaN             NaN          NaN

   title_year actor_2_facebook_likes imdb_score  aspect_ratio  \
0      2009.0                 936.0        7.9          1.78
1      2007.0                5000.0        7.1          2.35
2      2015.0                 393.0        6.8          2.35
3      2012.0               23000.0        8.5          2.35
4         NaN                  12.0        7.1           NaN

  movie_facebook_likes
0                33000
1                    0
2                85000
3               164000
4                    0

[5 rows x 28 columns]
```

[6]:
```
movies.shape
```

[6]:
```
(5043, 28)
```

[7]:
```
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5043 entries, 0 to 5042
Data columns (total 28 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   color                      5024 non-null   object
 1   director_name              4939 non-null   object
 2   num_critic_for_reviews     4993 non-null   float64
 3   duration                   5028 non-null   float64
 4   director_facebook_likes    4939 non-null   float64
 5   actor_3_facebook_likes     5020 non-null   float64
 6   actor_2_name               5030 non-null   object
 7   actor_1_facebook_likes     5036 non-null   float64
 8   gross                      4159 non-null   float64
 9   genres                     5043 non-null   object
 10  actor_1_name               5036 non-null   object
 11  movie_title                5043 non-null   object
 12  num_voted_users            5043 non-null   int64
```

```
 13  cast_total_facebook_likes  5043 non-null   int64
 14  actor_3_name               5020 non-null   object
 15  facenumber_in_poster       5030 non-null   float64
 16  plot_keywords              4890 non-null   object
 17  movie_imdb_link            5043 non-null   object
 18  num_user_for_reviews       5022 non-null   float64
 19  language                   5031 non-null   object
 20  country                    5038 non-null   object
 21  content_rating             4740 non-null   object
 22  budget                     4551 non-null   float64
 23  title_year                 4935 non-null   float64
 24  actor_2_facebook_likes     5030 non-null   float64
 25  imdb_score                 5043 non-null   float64
 26  aspect_ratio               4714 non-null   float64
 27  movie_facebook_likes       5043 non-null   int64
dtypes: float64(13), int64(3), object(12)
memory usage: 1.1+ MB
```

[8]:
```
movies.describe()
```

[8]:

	num_critic_for_reviews	duration	director_facebook_likes	\
count	4993.000000	5028.000000	4939.000000	
mean	140.194272	107.201074	686.509212	
std	121.601675	25.197441	2813.328607	
min	1.000000	7.000000	0.000000	
25%	50.000000	93.000000	7.000000	
50%	110.000000	103.000000	49.000000	
75%	195.000000	118.000000	194.500000	
max	813.000000	511.000000	23000.000000	

	actor_3_facebook_likes	actor_1_facebook_likes	gross	\
count	5020.000000	5036.000000	4.159000e+03	
mean	645.009761	6560.047061	4.846841e+07	
std	1665.041728	15020.759120	6.845299e+07	
min	0.000000	0.000000	1.620000e+02	
25%	133.000000	614.000000	5.340988e+06	
50%	371.500000	988.000000	2.551750e+07	
75%	636.000000	11000.000000	6.230944e+07	
max	23000.000000	640000.000000	7.605058e+08	

	num_voted_users	cast_total_facebook_likes	facenumber_in_poster	\
count	5.043000e+03	5043.000000	5030.000000	
mean	8.366816e+04	9699.063851	1.371173	
std	1.384853e+05	18163.799124	2.013576	
min	5.000000e+00	0.000000	0.000000	
25%	8.593500e+03	1411.000000	0.000000	
50%	3.435900e+04	3090.000000	1.000000	

```
75%       9.630900e+04              13756.500000             2.000000
max       1.689764e+06             656730.000000            43.000000

       num_user_for_reviews        budget   title_year  \
count           5022.000000  4.551000e+03  4935.000000
mean             272.770808  3.975262e+07  2002.470517
std              377.982886  2.061149e+08    12.474599
min                1.000000  2.180000e+02  1916.000000
25%               65.000000  6.000000e+06  1999.000000
50%              156.000000  2.000000e+07  2005.000000
75%              326.000000  4.500000e+07  2011.000000
max             5060.000000  1.221550e+10  2016.000000

       actor_2_facebook_likes   imdb_score  aspect_ratio  movie_facebook_likes
count             5030.000000  5043.000000   4714.000000           5043.000000
mean              1651.754473     6.442138      2.220403           7525.964505
std               4042.438863     1.125116      1.385113          19320.445110
min                  0.000000     1.600000      1.180000              0.000000
25%                281.000000     5.800000      1.850000              0.000000
50%                595.000000     6.600000      2.350000            166.000000
75%                918.000000     7.200000      2.350000           3000.000000
max             137000.000000     9.500000     16.000000         349000.000000
```

2 Task 2: Cleaning the Data

Subtask 2.1: Inspect Null values

Find out the number of Null values in all the columns and rows. Also, find the percentage of Null values in each column. Round-off the percentages upto two decimal places.

[9]:
```
movies.isnull().sum()
```

[9]:
```
color                       19
director_name              104
num_critic_for_reviews      50
duration                    15
director_facebook_likes    104
actor_3_facebook_likes      23
actor_2_name                13
actor_1_facebook_likes       7
gross                      884
genres                       0
actor_1_name                 7
movie_title                  0
num_voted_users              0
cast_total_facebook_likes    0
actor_3_name                23
```

```
facenumber_in_poster          13
plot_keywords                153
movie_imdb_link               0
num_user_for_reviews          21
language                      12
country                       5
content_rating               303
budget                       492
title_year                   108
actor_2_facebook_likes        13
imdb_score                    0
aspect_ratio                 329
movie_facebook_likes          0
dtype: int64
```

[10]:
```
movies.isnull().sum(axis=1)
```

[10]:
```
0        0
1        0
2        0
3        0
4       14
        ..
5038     4
5039     5
5040     4
5041     2
5042     0
Length: 5043, dtype: int64
```

[11]:
```
# Get the percentages by dividing the sum obtained previously by the total␣
 ↪length, multiplying it by 100 and rounding it off to
# two decimal places
round(100*(movies.isnull().sum()/len(movies.index)),2)
```

[11]:
```
color                         0.38
director_name                 2.06
num_critic_for_reviews        0.99
duration                      0.30
director_facebook_likes       2.06
actor_3_facebook_likes        0.46
actor_2_name                  0.26
actor_1_facebook_likes        0.14
gross                        17.53
genres                        0.00
actor_1_name                  0.14
movie_title                   0.00
```

```
num_voted_users               0.00
cast_total_facebook_likes     0.00
actor_3_name                  0.46
facenumber_in_poster          0.26
plot_keywords                 3.03
movie_imdb_link               0.00
num_user_for_reviews          0.42
language                      0.24
country                       0.10
content_rating                6.01
budget                        9.76
title_year                    2.14
actor_2_facebook_likes        0.26
imdb_score                    0.00
aspect_ratio                  6.52
movie_facebook_likes          0.00
dtype: float64
```

Subtask 2.2: Drop unecessary columns

For this assignment, you will mostly be analyzing the movies with respect to the ratings, gross collection, popularity of movies, etc. So many of the columns in this dataframe are not required. So it is advised to drop the following columns.

color,

director_facebook_likes,

actor_1_facebook_likes,

actor_2_facebook_likes,

actor_3_facebook_likes,

actor_2_name ,

cast_total_facebook_likes,

actor_3_name,

duration,

facenumber_in_poster,

content_rating,

country,

movie_imdb_link,

aspect_ratio,

plot_keywords,

[12]:
```
movies= movies.drop(['color',
                     'director_facebook_likes',
                     'actor_3_facebook_likes',
                     'actor_1_facebook_likes',
                     'cast_total_facebook_likes',
                     'actor_2_facebook_likes',
                     'duration',
                     'facenumber_in_poster',
                     'content_rating',
                     'country',
                     'movie_imdb_link',
                     'aspect_ratio',
                     'plot_keywords',
                     'actor_2_name',
                     'actor_3_name'],axis=1)
```

[13]:
```
movies.shape
```

[13]:
```
(5043, 13)
```

Subtask 2.3: Drop unecessary rows using columns with high NaN percentages

On inspection you might notice that some columns have large percentage (greater than 5%) of Null values. Drop all the rows which have Null values for such columns.

[14]:
```
# Inspecting the percentages of Null values again

round(100*(movies.isnull().sum()/len(movies.index)), 2)
```

[14]:
```
director_name              2.06
num_critic_for_reviews     0.99
gross                     17.53
genres                     0.00
actor_1_name               0.14
movie_title                0.00
num_voted_users            0.00
num_user_for_reviews       0.42
language                   0.24
budget                     9.76
title_year                 2.14
imdb_score                 0.00
movie_facebook_likes       0.00
dtype: float64
```

[15]:
```
# Since 'gross' and 'budget' columns have large number of NaN values, drop all␣
  ↪the rows with NaNs at this column using the
# 'isnan' function of NumPy alongwith a negation '~'
movies= movies[~np.isnan(movies['gross'])]
```

```
movies= movies[~np.isnan(movies['budget'])]
```

[16]:
```
# Inspecting the percentages of NaN

round(100*(movies.isnull().sum()/len(movies.index)), 2)
```

[16]:
```
director_name              0.00
num_critic_for_reviews     0.03
gross                      0.00
genres                     0.00
actor_1_name               0.08
movie_title                0.00
num_voted_users            0.00
num_user_for_reviews       0.00
language                   0.08
budget                     0.00
title_year                 0.00
imdb_score                 0.00
movie_facebook_likes       0.00
dtype: float64
```

Subtask 2.4: Drop unecessary rows

Some of the rows might have greater than five Null values. Such rows aren't of much use for the analysis and hence, should be removed.

[18]:
```
# The rows for which the sum of Null is less than five are retained

movies = movies[movies.isnull().sum(axis=1) <= 5]
movies
```

[18]:
```
          director_name  num_critic_for_reviews        gross  \
0         James Cameron                   723.0  760505847.0
1        Gore Verbinski                   302.0  309404152.0
2            Sam Mendes                   602.0  200074175.0
3     Christopher Nolan                   813.0  448130642.0
5        Andrew Stanton                   462.0   73058679.0
…                   …                       …           …
5033      Shane Carruth                   143.0     424760.0
5034   Neill Dela Llana                    35.0      70071.0
5035   Robert Rodriguez                    56.0    2040920.0
5037       Edward Burns                    14.0       4584.0
5042           Jon Gunn                    43.0      85222.0

                                  genres     actor_1_name  \
0         Action|Adventure|Fantasy|Sci-Fi      CCH Pounder
1                Action|Adventure|Fantasy      Johnny Depp
2               Action|Adventure|Thriller  Christoph Waltz
```

```
3                          Action|Thriller        Tom Hardy
5                 Action|Adventure|Sci-Fi     Daryl Sabara
…                                     …              …
5033                 Drama|Sci-Fi|Thriller    Shane Carruth
5034                               Thriller      Ian Gamazon
5035  Action|Crime|Drama|Romance|Thriller  Carlos Gallardo
5037                        Comedy|Drama      Kerry Bishé
5042                         Documentary      John August

                                    movie_title  num_voted_users  \
0                                        Avatar           886204
1     Pirates of the Caribbean: At World's End           471220
2                                      Spectre           275868
3                        The Dark Knight Rises          1144337
5                                   John Carter           212204
…                                           …              …
5033                                    Primer            72639
5034                                    Cavite              589
5035                                 El Mariachi            52055
5037                                   Newlyweds             1338
5042                           My Date with Drew             4285

      num_user_for_reviews language       budget  title_year  imdb_score  \
0                   3054.0  English  237000000.0      2009.0         7.9
1                   1238.0  English  300000000.0      2007.0         7.1
2                    994.0  English  245000000.0      2015.0         6.8
3                   2701.0  English  250000000.0      2012.0         8.5
5                    738.0  English  263700000.0      2012.0         6.6
…                      …      …           …          …          …
5033                 371.0  English       7000.0      2004.0         7.0
5034                  35.0  English       7000.0      2005.0         6.3
5035                 130.0  Spanish       7000.0      1992.0         6.9
5037                  14.0  English       9000.0      2011.0         6.4
5042                  84.0  English       1100.0      2004.0         6.6

      movie_facebook_likes
0                    33000
1                        0
2                    85000
3                   164000
5                    24000
…                        …
5033                 19000
5034                    74
5035                     0
5037                   413
5042                   456
```

```
[3891 rows x 13 columns]
```

[19]:
```
# Inspecting the percentages of NaN

round(100*(movies.isnull().sum()/len(movies.index)), 2)
```

[19]:
```
director_name              0.00
num_critic_for_reviews     0.03
gross                      0.00
genres                     0.00
actor_1_name               0.08
movie_title                0.00
num_voted_users            0.00
num_user_for_reviews       0.00
language                   0.08
budget                     0.00
title_year                 0.00
imdb_score                 0.00
movie_facebook_likes       0.00
dtype: float64
```

Subtask 2.5: Fill NaN values

You might notice that the language column has some NaN values. Here, on inspection, you will see that it is safe to replace all the missing values with 'English'.

[20]:
```
# Inspect the language column of the dataset
movies['language'].describe()
```

[20]:
```
count        3888
unique         38
top       English
freq         3707
Name: language, dtype: object
```

[22]:
```
movies['language'].mode()[0]
```

[22]:
```
'English'
```

[23]:
```
movies['language']= movies['language'].fillna(movies['language'].mode()[0])
```

[24]:
```
round(100*(movies.isnull().sum()/len(movies.index)), 2)
```

[24]:
```
director_name              0.00
num_critic_for_reviews     0.03
gross                      0.00
genres                     0.00
```

```
actor_1_name               0.08
movie_title                0.00
num_voted_users            0.00
num_user_for_reviews       0.00
language                   0.00
budget                     0.00
title_year                 0.00
imdb_score                 0.00
movie_facebook_likes       0.00
dtype: float64
```

Subtask 2.6: Check the number of retained rows

You might notice that two of the columns viz. num_critic_for_reviews and actor_1_name have small percentages of NaN values left. You can let these columns as it is for now. Check the number and percentage of the rows retained after completing all the tasks above.

[25]:
```
# Get the number of retained rows using 'len()'
# Get the percentage of retained rows by dividing the current number of rows␣
 ↪with initial number of rows

print(len(movies.index))
print(len(movies.index)/5042)
```

```
3891
0.771717572391908
```

3 Task 3: Data Analysis

Subtask 3.1: Change the unit of column Convert the unit of the budget and gross columns from ''*tomillion*.

[26]:
```
# Divide the 'gross' and 'budget' columns by 1000000 to convert '$' to 'million␣
 ↪$'
movies['gross'] = movies['gross']/1000000
movies['budget'] = movies['budget']/1000000
movies
```

[26]:

	director_name	num_critic_for_reviews	gross	\
0	James Cameron	723.0	760.505847	
1	Gore Verbinski	302.0	309.404152	
2	Sam Mendes	602.0	200.074175	
3	Christopher Nolan	813.0	448.130642	
5	Andrew Stanton	462.0	73.058679	
…	…	…	…	
5033	Shane Carruth	143.0	0.424760	
5034	Neill Dela Llana	35.0	0.070071	
5035	Robert Rodriguez	56.0	2.040920	

```
5037       Edward Burns                      14.0    0.004584
5042           Jon Gunn                      43.0    0.085222

                                    genres     actor_1_name  \
0         Action|Adventure|Fantasy|Sci-Fi      CCH Pounder
1                Action|Adventure|Fantasy      Johnny Depp
2               Action|Adventure|Thriller  Christoph Waltz
3                         Action|Thriller        Tom Hardy
5                  Action|Adventure|Sci-Fi     Daryl Sabara
…                                      …             …
5033                Drama|Sci-Fi|Thriller    Shane Carruth
5034                              Thriller      Ian Gamazon
5035  Action|Crime|Drama|Romance|Thriller  Carlos Gallardo
5037                        Comedy|Drama      Kerry Bishé
5042                         Documentary      John August

                                    movie_title  num_voted_users  \
0                                        Avatar           886204
1     Pirates of the Caribbean: At World's End           471220
2                                       Spectre           275868
3                         The Dark Knight Rises          1144337
5                                   John Carter           212204
…                                           …              …
5033                                    Primer            72639
5034                                    Cavite              589
5035                                 El Mariachi            52055
5037                                   Newlyweds             1338
5042                           My Date with Drew             4285

      num_user_for_reviews language    budget  title_year  imdb_score  \
0                   3054.0  English  237.0000      2009.0         7.9
1                   1238.0  English  300.0000      2007.0         7.1
2                    994.0  English  245.0000      2015.0         6.8
3                   2701.0  English  250.0000      2012.0         8.5
5                    738.0  English  263.7000      2012.0         6.6
…                      …      …       …          …          …
5033                 371.0  English    0.0070      2004.0         7.0
5034                  35.0  English    0.0070      2005.0         6.3
5035                 130.0  Spanish    0.0070      1992.0         6.9
5037                  14.0  English    0.0090      2011.0         6.4
5042                  84.0  English    0.0011      2004.0         6.6

      movie_facebook_likes
0                    33000
1                        0
2                    85000
3                   164000
```

```
5                    24000
…                    …
5033                 19000
5034                    74
5035                     0
5037                   413
5042                   456

[3891 rows x 13 columns]
```

Subtask 3.2: Find the movies with highest 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

[27]:
```
# Create the new column named 'profit' by subtracting the 'budget' column from␣
↪the 'gross' column
movies['profit']= movies['gross'] - movies['budget']
movies.head()
```

[27]:
```
       director_name  num_critic_for_reviews       gross  \
0      James Cameron                   723.0  760.505847
1     Gore Verbinski                   302.0  309.404152
2         Sam Mendes                   602.0  200.074175
3  Christopher Nolan                   813.0  448.130642
5     Andrew Stanton                   462.0   73.058679

                            genres     actor_1_name  \
0  Action|Adventure|Fantasy|Sci-Fi      CCH Pounder
1         Action|Adventure|Fantasy      Johnny Depp
2        Action|Adventure|Thriller  Christoph Waltz
3                  Action|Thriller        Tom Hardy
5          Action|Adventure|Sci-Fi     Daryl Sabara

                                movie_title  num_voted_users  \
0                                    Avatar           886204
1  Pirates of the Caribbean: At World's End           471220
2                                   Spectre           275868
3                    The Dark Knight Rises          1144337
5                               John Carter           212204

   num_user_for_reviews language  budget  title_year  imdb_score  \
0                3054.0  English   237.0      2009.0         7.9
1                1238.0  English   300.0      2007.0         7.1
```

```
2                994.0  English   245.0      2015.0         6.8
3               2701.0  English   250.0      2012.0         8.5
5                738.0  English   263.7      2012.0         6.6

   movie_facebook_likes      profit
0                 33000  523.505847
1                     0    9.404152
2                 85000  -44.925825
3                164000  198.130642
5                 24000 -190.641321
```

[28]:
```
# Sort the dataframe with the 'profit' column as reference using the␣
 ↪'sort_values' function. Make sure to set the argument
# 'ascending' to 'False'

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

[28]:
```
          director_name  num_critic_for_reviews       gross  \
0         James Cameron                   723.0  760.505847
29      Colin Trevorrow                   644.0  652.177271
26        James Cameron                   315.0  658.672302
3024       George Lucas                   282.0  460.935665
3080   Steven Spielberg                   215.0  434.949459
…                  …                       …           …
2334    Katsuhiro Ôtomo                   105.0    0.410388
2323     Hayao Miyazaki                   174.0    2.298191
3005       Lajos Koltai                    73.0    0.195888
3859     Chan-wook Park                   202.0    0.211667
2988       Joon-ho Bong                   363.0    2.201412

                                              genres         actor_1_name  \
0                        Action|Adventure|Fantasy|Sci-Fi          CCH Pounder
29                     Action|Adventure|Sci-Fi|Thriller  Bryce Dallas Howard
26                                     Drama|Romance    Leonardo DiCaprio
3024                    Action|Adventure|Fantasy|Sci-Fi        Harrison Ford
3080                                    Family|Sci-Fi         Henry Thomas
…                                                 …                  …
2334  Action|Adventure|Animation|Family|Sci-Fi|Thriller     William Hootkins
2323                        Adventure|Animation|Fantasy        Minnie Driver
3005                               Drama|Romance|War         Marcell Nagy
3859                                     Crime|Drama         Min-sik Choi
2988                       Comedy|Drama|Horror|Sci-Fi            Doona Bae

                            movie_title  num_voted_users  \
0                                Avatar           886204
29                      Jurassic World           418214
```

```
26                              Titanic            793059
3024  Star Wars: Episode IV - A New Hope            911097
3080         E.T. the Extra-Terrestrial            281842
…                                    …                 …
2334                            Steamboy             13727
2323                  Princess Mononoke            221552
3005                           Fateless              5603
3859                      Lady Vengeance             53508
2988                           The Host              68883

      num_user_for_reviews   language        budget  title_year  imdb_score  \
0                   3054.0    English    237.000000      2009.0         7.9
29                  1290.0    English    150.000000      2015.0         7.0
26                  2528.0    English    200.000000      1997.0         7.7
3024                1470.0    English     11.000000      1977.0         8.7
3080                 515.0    English     10.500000      1982.0         7.9
…                      …          …             …           …           …
2334                  79.0   Japanese   2127.519898      2004.0         6.9
2323                 570.0   Japanese   2400.000000      1997.0         8.4
3005                  45.0  Hungarian   2500.000000      2005.0         7.1
3859                 131.0     Korean   4200.000000      2005.0         7.7
2988                 279.0     Korean  12215.500000      2006.0         7.0

      movie_facebook_likes        profit
0                    33000    523.505847
29                  150000    502.177271
26                   26000    458.672302
3024                 33000    449.935665
3080                 34000    424.449459
…                      …           …
2334                   973  -2127.109510
2323                 11000  -2397.701809
3005                   607  -2499.804112
3859                  4000  -4199.788333
2988                  7000 -12213.298588

[3891 rows x 14 columns]
```

[29]:
```
# Get the top 10 profitable movies by using position based indexing. Specify␣
↪the rows till 10 (0-9)

top10 = movies.iloc[:10, ]
top10
```

[29]:
```
          director_name  num_critic_for_reviews       gross  \
0         James Cameron                   723.0  760.505847
29      Colin Trevorrow                   644.0  652.177271
```

```
26        James Cameron                   315.0  658.672302
3024       George Lucas                   282.0  460.935665
3080   Steven Spielberg                   215.0  434.949459
794          Joss Whedon                  703.0  623.279547
17           Joss Whedon                  703.0  623.279547
509         Roger Allers                  186.0  422.783777
240         George Lucas                  320.0  474.544677
66     Christopher Nolan                  645.0  533.316061

                                      genres         actor_1_name  \
0               Action|Adventure|Fantasy|Sci-Fi          CCH Pounder
29            Action|Adventure|Sci-Fi|Thriller  Bryce Dallas Howard
26                               Drama|Romance    Leonardo DiCaprio
3024            Action|Adventure|Fantasy|Sci-Fi        Harrison Ford
3080                             Family|Sci-Fi         Henry Thomas
794                    Action|Adventure|Sci-Fi      Chris Hemsworth
17                     Action|Adventure|Sci-Fi      Chris Hemsworth
509   Adventure|Animation|Drama|Family|Musical    Matthew Broderick
240             Action|Adventure|Fantasy|Sci-Fi      Natalie Portman
66                  Action|Crime|Drama|Thriller       Christian Bale

                                    movie_title  num_voted_users  \
0                                        Avatar           886204
29                               Jurassic World           418214
26                                      Titanic           793059
3024         Star Wars: Episode IV - A New Hope           911097
3080                 E.T. the Extra-Terrestrial           281842
794                                The Avengers           995415
17                                 The Avengers           995415
509                               The Lion King           644348
240   Star Wars: Episode I - The Phantom Menace           534658
66                               The Dark Knight          1676169

      num_user_for_reviews language  budget  title_year  imdb_score  \
0                   3054.0  English   237.0      2009.0         7.9
29                  1290.0  English   150.0      2015.0         7.0
26                  2528.0  English   200.0      1997.0         7.7
3024                1470.0  English    11.0      1977.0         8.7
3080                 515.0  English    10.5      1982.0         7.9
794                 1722.0  English   220.0      2012.0         8.1
17                  1722.0  English   220.0      2012.0         8.1
509                  656.0  English    45.0      1994.0         8.5
240                 3597.0  English   115.0      1999.0         6.5
66                  4667.0  English   185.0      2008.0         9.0

      movie_facebook_likes      profit
0                    33000  523.505847
```

```
29                  150000  502.177271
26                   26000  458.672302
3024                 33000  449.935665
3080                 34000  424.449459
794                 123000  403.279547
17                  123000  403.279547
509                  17000  377.783777
240                  13000  359.544677
66                   37000  348.316061
```

Subtask 3.3: Drop duplicate values

After you found out the top 10 profiting movies, you might have noticed a duplicate value. So, it seems like the dataframe has duplicate values as well. Drop the duplicate values from the dataframe and repeat Subtask 3.2.

[30]:
```
# Drop the duplicate values using 'drop_duplicates' function. All the columns␣
 ↪for duplicate rows need to be dropped and thus,
# the 'subset' argument is set to 'None'. The 'keep = first' indicates to␣
 ↪retain the first row among the duplicate rows, and
# 'inplace = True' performs the operation on the dataframe in place.

movies.drop_duplicates(subset = None, keep = 'first', inplace = True)
movies
```

[30]:
```
          director_name  num_critic_for_reviews       gross  \
0         James Cameron                   723.0  760.505847
29      Colin Trevorrow                   644.0  652.177271
26        James Cameron                   315.0  658.672302
3024       George Lucas                   282.0  460.935665
3080   Steven Spielberg                   215.0  434.949459
…                   …                     …           …
2334    Katsuhiro Ôtomo                   105.0    0.410388
2323     Hayao Miyazaki                   174.0    2.298191
3005       Lajos Koltai                    73.0    0.195888
3859     Chan-wook Park                   202.0    0.211667
2988       Joon-ho Bong                   363.0    2.201412

                                                genres         actor_1_name  \
0                         Action|Adventure|Fantasy|Sci-Fi          CCH Pounder
29                      Action|Adventure|Sci-Fi|Thriller  Bryce Dallas Howard
26                                    Drama|Romance    Leonardo DiCaprio
3024                    Action|Adventure|Fantasy|Sci-Fi        Harrison Ford
3080                                  Family|Sci-Fi         Henry Thomas
…                                                 …                  …
2334  Action|Adventure|Animation|Family|Sci-Fi|Thriller     William Hootkins
2323                          Adventure|Animation|Fantasy        Minnie Driver
3005                                  Drama|Romance|War         Marcell Nagy
```

```
3859                                      Crime|Drama         Min-sik Choi
2988                          Comedy|Drama|Horror|Sci-Fi            Doona Bae

                            movie_title  num_voted_users  \
0                                Avatar           886204
29                      Jurassic World           418214
26                             Titanic           793059
3024  Star Wars: Episode IV - A New Hope           911097
3080          E.T. the Extra-Terrestrial           281842
…                                     …              …
2334                          Steamboy            13727
2323                 Princess Mononoke           221552
3005                          Fateless             5603
3859                     Lady Vengeance            53508
2988                          The Host             68883

      num_user_for_reviews   language        budget  title_year  imdb_score  \
0                   3054.0    English    237.000000      2009.0         7.9
29                  1290.0    English    150.000000      2015.0         7.0
26                  2528.0    English    200.000000      1997.0         7.7
3024                1470.0    English     11.000000      1977.0         8.7
3080                 515.0    English     10.500000      1982.0         7.9
…                      …          …             …           …           …
2334                  79.0   Japanese   2127.519898      2004.0         6.9
2323                 570.0   Japanese   2400.000000      1997.0         8.4
3005                  45.0  Hungarian   2500.000000      2005.0         7.1
3859                 131.0     Korean   4200.000000      2005.0         7.7
2988                 279.0     Korean  12215.500000      2006.0         7.0

      movie_facebook_likes        profit
0                    33000    523.505847
29                  150000    502.177271
26                   26000    458.672302
3024                 33000    449.935665
3080                 34000    424.449459
…                      …             …
2334                   973  -2127.109510
2323                 11000  -2397.701809
3005                   607  -2499.804112
3859                  4000  -4199.788333
2988                  7000 -12213.298588

[3856 rows x 14 columns]
```

[31]:
```
# Get the top 10 profitable movies by using position based indexing. Specify␣
↪the rows till 10 (0-9)
```

```
top10 = movies.iloc[:10, ]
top10
```

[31]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_for_reviews	language	budget	title_year	imdb_score
0	James Cameron	723.0	760.505847	Action|Adventure|Fantasy|Sci-Fi	CCH Pounder	Avatar	886204	3054.0	English	237.0	2009.0	7.9
29	Colin Trevorrow	644.0	652.177271	Action|Adventure|Sci-Fi|Thriller	Bryce Dallas Howard	Jurassic World	418214	1290.0	English	150.0	2015.0	7.0
26	James Cameron	315.0	658.672302	Drama|Romance	Leonardo DiCaprio	Titanic	793059	2528.0	English	200.0	1997.0	7.7
3024	George Lucas	282.0	460.935665	Action|Adventure|Fantasy|Sci-Fi	Harrison Ford	Star Wars: Episode IV - A New Hope	911097	1470.0	English	11.0	1977.0	8.7
3080	Steven Spielberg	215.0	434.949459	Family|Sci-Fi	Henry Thomas	E.T. the Extra-Terrestrial	281842	515.0	English	10.5	1982.0	7.9
794	Joss Whedon	703.0	623.279547	Action|Adventure|Sci-Fi	Chris Hemsworth	The Avengers	995415	1722.0	English	220.0	2012.0	8.1
509	Roger Allers	186.0	422.783777	Adventure|Animation|Drama|Family|Musical	Matthew Broderick	The Lion King	644348	656.0	English	45.0	1994.0	8.5
240	George Lucas	320.0	474.544677	Action|Adventure|Fantasy|Sci-Fi	Natalie Portman	Star Wars: Episode I - The Phantom Menace	534658					
66	Christopher Nolan	645.0	533.316061	Action|Crime|Drama|Thriller	Christian Bale	The Dark Knight	1676169					
439	Gary Ross	673.0	407.999255	Adventure|Drama|Sci-Fi|Thriller	Jennifer Lawrence	The Hunger Games	701607					

```
240                3597.0  English   115.0      1999.0         6.5
66                 4667.0  English   185.0      2008.0         9.0
439                1959.0  English    78.0      2012.0         7.3

      movie_facebook_likes      profit
0                    33000  523.505847
29                  150000  502.177271
26                   26000  458.672302
3024                 33000  449.935665
3080                 34000  424.449459
794                 123000  403.279547
509                  17000  377.783777
240                  13000  359.544677
66                   37000  348.316061
439                 140000  329.999255
```

Subtask 3.4: Find IMDb Top 250

1.Create a new dataframe IMDb_Top_250 and store the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb_score). Also make sure that for all of these movies, the num_voted_users is greater than 25,000. Also add a Rank column containing the values 1 to 250 indicating the ranks of the corresponding films.

2.Extract all the movies in the IMDb_Top_250 dataframe which are not in the English language and store them in a new dataframe named Top_Foreign_Lang_Film.‘

[32]:
```
# Sort the movies by IMDb score
# Retain the movies with 'num_voted_users' greater than 25000
# Use position based indexing to get the first 250 rows in the sorted dataframe
# Create a new column rank which contains the rank from 1 to 250

IMDb_Top_250 = movies.sort_values(by='imdb_score', ascending= False)
IMDb_Top_250 = IMDb_Top_250.loc[IMDb_Top_250.num_voted_users>25000]
IMDb_Top_250 = IMDb_Top_250.iloc[:250, ]
IMDb_Top_250['Rank'] = range(1,251)
IMDb_Top_250
```

[32]:
```
             director_name  num_critic_for_reviews       gross  \
1937        Frank Darabont                   199.0   28.341469
3466  Francis Ford Coppola                   208.0  134.821952
66       Christopher Nolan                   645.0  533.316061
2837  Francis Ford Coppola                   149.0   57.300000
339          Peter Jackson                   328.0  377.019252
…                      …                       …         …
788          Cameron Crowe                   149.0   32.522352
99           Peter Jackson                   645.0  303.001229
1606       Nick Cassavetes                   177.0    0.064286
1735         James Mangold                   291.0  119.518352
```

```
639          Michael Mann                   209.0   28.965197

                             genres            actor_1_name  \
1937                    Crime|Drama          Morgan Freeman
3466                    Crime|Drama               Al Pacino
66       Action|Crime|Drama|Thriller          Christian Bale
2837                    Crime|Drama          Robert De Niro
339   Action|Adventure|Drama|Fantasy           Orlando Bloom
…                                 …                     …
788     Adventure|Comedy|Drama|Music  Philip Seymour Hoffman
99                 Adventure|Fantasy            Aidan Turner
1606                   Drama|Romance            Ryan Gosling
1735   Biography|Drama|Music|Romance   Sandra Ellis Lafferty
639         Biography|Drama|Thriller               Al Pacino

                                       movie_title  num_voted_users  \
1937                      The Shawshank Redemption          1689764
3466                                 The Godfather          1155770
66                                 The Dark Knight          1676169
2837                         The Godfather: Part II           790926
339   The Lord of the Rings: The Return of the King          1215718
…                                               …                …
788                                  Almost Famous           207287
99               The Hobbit: An Unexpected Journey           637246
1606                                The Notebook           396396
1735                               Walk the Line           188637
639                                  The Insider           133526

      num_user_for_reviews language  budget  title_year  imdb_score  \
1937                4144.0  English    25.0      1994.0         9.3
3466                2238.0  English     6.0      1972.0         9.2
66                  4667.0  English   185.0      2008.0         9.0
2837                 650.0  English    13.0      1974.0         9.0
339                 3189.0  English    94.0      2003.0         8.9
…                      …      …     …          …          …
788                  822.0  English    60.0      2000.0         7.9
99                  1367.0  English   180.0      2012.0         7.9
1606                1111.0  English    29.0      2004.0         7.9
1735                 815.0  English    28.0      2005.0         7.9
639                  521.0  English    68.0      1999.0         7.9

      movie_facebook_likes      profit  Rank
1937                108000    3.341469     1
3466                 43000  128.821952     2
66                   37000  348.316061     3
2837                 14000   44.300000     4
339                  16000  283.019252     5
```

```
…                    …          …   …
788                  15000  -27.477648   246
99                  166000  123.001229   247
1606                 57000  -28.935714   248
1735                 11000   91.518352   249
639                      0  -39.034803   250

[250 rows x 15 columns]
```

[33]:
```
# Get the non-English language films using conditional label based indexing

Top_Foreign_Lang_Film = IMDb_Top_250.loc[IMDb_Top_250['language'] != 'English']
Top_Foreign_Lang_Film
```

[33]:

	director_name	num_critic_for_reviews	gross	\
4498	Sergio Leone	181.0	6.100000	
4029	Fernando Meirelles	214.0	7.563397	
4747	Akira Kurosawa	153.0	0.269061	
2373	Hayao Miyazaki	246.0	10.049886	
4259	Florian Henckel von Donnersmarck	215.0	11.284657	
4921	Majid Majidi	46.0	0.925402	
4105	Chan-wook Park	305.0	2.181290	
1298	Jean-Pierre Jeunet	242.0	33.201661	
1329	S.S. Rajamouli	44.0	6.498000	
2323	Hayao Miyazaki	174.0	2.298191	
2970	Wolfgang Petersen	96.0	11.433134	
4659	Asghar Farhadi	354.0	7.098492	
4033	Thomas Vinterberg	349.0	0.610968	
2829	Oliver Hirschbiegel	192.0	5.501940	
2734	Fritz Lang	260.0	0.026435	
3550	Denis Villeneuve	226.0	6.857096	
4000	Juan José Campanella	262.0	20.167424	
2047	Hayao Miyazaki	212.0	4.710455	
2551	Guillermo del Toro	406.0	37.623143	
3553	José Padilha	142.0	0.008060	
2914	Je-kyu Kang	86.0	1.110186	
2830	Alejandro Amenábar	157.0	2.086345	
4267	Alejandro G. Iñárritu	157.0	5.383834	
3423	Katsuhiro Ôtomo	150.0	0.439162	
4461	Thomas Vinterberg	98.0	1.647780	
3344	Karan Johar	210.0	4.018695	
4284	Ari Folman	231.0	2.283276	
3456	Vincent Paronnaud	242.0	4.443403	
4144	Walter Salles	71.0	5.595428	
4897	Sergio Leone	122.0	3.500000	
3677	Christophe Barratier	112.0	3.629758	
4640	Cristian Mungiu	233.0	1.185783	

```
4415                  Fabián Bielinsky                    94.0   1.221261
2863                    Clint Eastwood                   251.0  13.753931
3510                       Yash Chopra                    29.0   2.921738
3264                    Michael Haneke                   447.0   0.225377

                                                 genres  \
4498                                            Western
4029                                        Crime|Drama
4747                           Action|Adventure|Drama
2373                  Adventure|Animation|Family|Fantasy
4259                                Drama|Thriller
4921                                 Drama|Family
4105                           Drama|Mystery|Thriller
1298                                Comedy|Romance
1329                  Action|Adventure|Drama|Fantasy|War
2323                         Adventure|Animation|Fantasy
2970                        Adventure|Drama|Thriller|War
4659                                Drama|Mystery
4033                                         Drama
2829                         Biography|Drama|History|War
2734                                 Drama|Sci-Fi
3550                              Drama|Mystery|War
4000                           Drama|Mystery|Thriller
2047                  Adventure|Animation|Family|Fantasy
2551                                Drama|Fantasy|War
3553                         Action|Crime|Drama|Thriller
2914                                  Action|Drama|War
2830                             Biography|Drama|Romance
4267                                 Drama|Thriller
3423                             Action|Animation|Sci-Fi
4461                                         Drama
3344                            Adventure|Drama|Thriller
4284  Animation|Biography|Documentary|Drama|History|War
3456                       Animation|Biography|Drama|War
4144                                         Drama
4897                               Action|Drama|Western
3677                                  Drama|Music
4640                                         Drama
4415                               Crime|Drama|Thriller
2863                                  Drama|History|War
3510                              Drama|Musical|Romance
3264                                  Drama|Romance

              actor_1_name                          movie_title  \
4498         Clint Eastwood     The Good, the Bad and the Ugly
4029            Alice Braga                       City of God
4747        Takashi Shimura                     Seven Samurai
```

```
2373         Bunta Sugawara                        Spirited Away
4259         Sebastian Koch                  The Lives of Others
4921         Bahare Seddiqi                   Children of Heaven
4105           Min-sik Choi                                Oldboy
1298      Mathieu Kassovitz                                Amélie
1329       Tamannaah Bhatia             Baahubali: The Beginning
2323          Minnie Driver                    Princess Mononoke
2970        Jürgen Prochnow                              Das Boot
4659        Shahab Hosseini                         A Separation
4033       Thomas Bo Larsen                             The Hunt
2829     Thomas Kretschmann                             Downfall
2734          Brigitte Helm                            Metropolis
3550           Lubna Azabal                             Incendies
4000          Ricardo Darín             The Secret in Their Eyes
2047         Christian Bale                 Howl's Moving Castle
2551          Ivana Baquero                      Pan's Labyrinth
3553           Wagner Moura                           Elite Squad
2914           Min-sik Choi  Tae Guk Gi: The Brotherhood of War
2830            Belén Rueda                       The Sea Inside
4267        Adriana Barraza                         Amores Perros
3423           Mitsuo Iwata                                 Akira
4461         Ulrich Thomsen                       The Celebration
3344         Shah Rukh Khan                       My Name Is Khan
4284             Ari Folman                    Waltz with Bashir
3456      Catherine Deneuve                             Persepolis
4144    Fernanda Montenegro                       Central Station
4897         Clint Eastwood                  A Fistful of Dollars
3677  Jean-Baptiste Maunier                             The Chorus
4640       Anamaria Marinca          4 Months, 3 Weeks and 2 Days
4415          Ricardo Darín                           Nine Queens
2863         Yuki Matsuzaki                Letters from Iwo Jima
3510         Shah Rukh Khan                             Veer-Zaara
3264       Isabelle Huppert                                  Amour

      num_voted_users  num_user_for_reviews    language       budget  \
4498           503509                 780.0     Italian     1.200000
4029           533200                 749.0  Portuguese     3.300000
4747           229012                 596.0    Japanese     2.000000
2373           417971                 902.0    Japanese    19.000000
4259           259379                 407.0      German     2.000000
4921            27882                 130.0     Persian     0.180000
4105           356181                 809.0      Korean     3.000000
1298           534262                1314.0      French    77.000000
1329            62756                 410.0      Telugu    18.026148
2323           221552                 570.0    Japanese  2400.000000
2970           168203                 426.0      German    14.000000
4659           151812                 264.0     Persian     0.500000
```

4033	170155	249.0	Danish	3.800000	
2829	248354	564.0	German	13.500000	
2734	111841	413.0	German	6.000000	
3550	80429	156.0	French	6.800000	
4000	131831	231.0	Spanish	2.000000	
2047	214091	330.0	Japanese	24.000000	
2551	467234	1083.0	Spanish	13.500000	
3553	81644	107.0	Portuguese	4.000000	
2914	31943	224.0	Korean	12.800000	
2830	64556	140.0	Spanish	10.000000	
4267	173551	361.0	Spanish	2.000000	
3423	106160	430.0	Japanese	1100.000000	
4461	65951	258.0	Danish	1.300000	
3344	69759	235.0	Hindi	12.000000	
4284	46107	156.0	Hebrew	1.500000	
3456	70194	158.0	French	7.300000	
4144	28951	257.0	Portuguese	2.900000	
4897	147566	235.0	Italian	0.200000	
3677	44151	110.0	French	5.500000	
4640	44763	172.0	Romanian	0.590000	
4415	38215	125.0	Spanish	1.500000	
2863	132149	316.0	Japanese	19.000000	
3510	34449	119.0	Hindi	7.000000	
3264	70382	190.0	French	8.900000	

	title_year	imdb_score	movie_facebook_likes	profit	Rank
4498	1966.0	8.9	20000	4.900000	7
4029	2002.0	8.7	28000	4.263397	17
4747	1954.0	8.7	11000	-1.730939	19
2373	2001.0	8.6	28000	-8.950114	22
4259	2006.0	8.5	39000	9.284657	29
4921	1997.0	8.5	0	0.745402	32
4105	2003.0	8.4	43000	-0.818710	48
1298	2001.0	8.4	39000	-43.798339	54
1329	2015.0	8.4	21000	-11.528148	57
2323	1997.0	8.4	11000	-2397.701809	58
2970	1981.0	8.4	11000	-2.566866	59
4659	2011.0	8.4	48000	6.598492	60
4033	2012.0	8.3	60000	-3.189032	63
2829	2004.0	8.3	14000	-7.998060	78
2734	1927.0	8.3	12000	-5.973565	84
3550	2010.0	8.2	37000	0.057096	91
4000	2009.0	8.2	33000	18.167424	92
2047	2004.0	8.2	13000	-19.289545	94
2551	2006.0	8.2	27000	24.123143	102
3553	2007.0	8.1	11000	-3.991940	108
2914	2004.0	8.1	0	-11.689814	121

2830	2004.0	8.1	0	-7.913655	124
4267	2000.0	8.1	11000	3.383834	132
3423	1988.0	8.1	0	-1099.560838	138
4461	1998.0	8.1	5000	0.347780	139
3344	2010.0	8.0	27000	-7.981305	157
4284	2008.0	8.0	0	0.783276	183
3456	2007.0	8.0	14000	-2.856597	184
4144	1998.0	8.0	0	2.695428	193
4897	1964.0	8.0	0	3.300000	200
3677	2004.0	7.9	0	-1.870242	206
4640	2007.0	7.9	14000	0.595783	208
4415	2000.0	7.9	0	-0.278739	215
2863	2006.0	7.9	5000	-5.246069	221
3510	2004.0	7.9	2000	-4.078262	228
3264	2012.0	7.9	33000	-8.674623	230

Subtask 3.5: Find the best directors

1. Group the dataframe using the director_name column. ‘
2. Find out the top 10 directors for whom the mean of imdb_score is the highest and store them in a new dataframe top10director.

[36]:
```
# Create a pivot table using 'director_name' as index, 'imdb_score' as values,␣
 ↪and 'mean' as aggfunc
# Sort the values by 'imdb_score'. Keep 'ascending' as 'False'
# Extract the top 10 from the dataframe created

# PS: If I had to find the worst 10 directors, I would have sorted the␣
 ↪dataframe in an ascending order and again extrcated the first 10 rows

director = movies.pivot_table(values = 'imdb_score', index = 'director_name',␣
 ↪aggfunc = 'mean')
director = director.sort_values(by = 'imdb_score', ascending = False)
director = director.iloc[:10, ]
director
```

[36]:

director_name	imdb_score
Charles Chaplin	8.600000
Tony Kaye	8.600000
Alfred Hitchcock	8.500000
Ron Fricke	8.500000
Damien Chazelle	8.500000
Majid Majidi	8.500000
Sergio Leone	8.433333
Christopher Nolan	8.425000
S.S. Rajamouli	8.400000
Marius A. Markevicius	8.400000

Subtask 3.6: Find popular genres

You might have noticed the genres column in the dataframe with all the genres of the movies seperated by a pipe (|). Out of all the movie genres, the first two are most significant for any film.

1. Extract the first two genres from the genres column and store them in two new columns: genre_1 and genre_2. Some of the movies might have only one genre. In such cases, extract the single genre into both the columns, i.e. for such movies the genre_2 should be the same as genre_1.
2. Group the dataframe using genre_1 as the primary column and genre_2 as the secondary column.
3. Find out the 5 most popular combo of genres by finding the mean of the gross values using the gross column and store them in a new dataframe named PopGenre.

[37]:
```
# Split the elements of the 'genre' column at the pipe characters ('|') using␣
 ↪str.split()
# Assign the first elements of the rows of 'genre' column to a new column named␣
 ↪'genre_1' using 'apply()' and 'lambda' functions
# Some of the movies have only one genre. In such cases, assign the same genre␣
 ↪to 'genre_2' as well

movies['genres'] = movies['genres'].str.split('|')
movies['genre_1'] = movies['genres'].apply(lambda x: x[0])
movies['genre_2'] = movies['genres'].apply(lambda x : x[1] if len(x) > 1 else␣
 ↪x[0])
movies
```

[37]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name
0	James Cameron	723.0	760.505847	[Action, Adventure, Fantasy, Sci-Fi]	CCH Pounder
29	Colin Trevorrow	644.0	652.177271	[Action, Adventure, Sci-Fi, Thriller]	Bryce Dallas Howard
26	James Cameron	315.0	658.672302	[Drama, Romance]	Leonardo DiCaprio
3024	George Lucas	282.0	460.935665	[Action, Adventure, Fantasy, Sci-Fi]	Harrison Ford
3080	Steven Spielberg	215.0	434.949459	[Family, Sci-Fi]	Henry Thomas
…	…	…	…	…	…
2334	Katsuhiro Ōtomo	105.0	0.410388	[Action, Adventure, Animation, Family, Sci-Fi,…	William Hootkins
2323	Hayao Miyazaki	174.0	2.298191		
3005	Lajos Koltai	73.0	0.195888		
3859	Chan-wook Park	202.0	0.211667		
2988	Joon-ho Bong	363.0	2.201412		

```
2323                    [Adventure, Animation, Fantasy]        Minnie Driver
3005                              [Drama, Romance, War]         Marcell Nagy
3859                                    [Crime, Drama]         Min-sik Choi
2988                    [Comedy, Drama, Horror, Sci-Fi]            Doona Bae

                            movie_title  num_voted_users  \
0                                Avatar           886204
29                      Jurassic World            418214
26                             Titanic            793059
3024  Star Wars: Episode IV - A New Hope           911097
3080         E.T. the Extra-Terrestrial            281842
…                                    …              …
2334                          Steamboy             13727
2323                 Princess Mononoke            221552
3005                         Fateless               5603
3859                    Lady Vengeance              53508
2988                         The Host               68883

      num_user_for_reviews   language        budget  title_year  imdb_score  \
0                   3054.0    English    237.000000      2009.0         7.9
29                  1290.0    English    150.000000      2015.0         7.0
26                  2528.0    English    200.000000      1997.0         7.7
3024                1470.0    English     11.000000      1977.0         8.7
3080                 515.0    English     10.500000      1982.0         7.9
…                      …          …             …           …           …
2334                  79.0   Japanese   2127.519898      2004.0         6.9
2323                 570.0   Japanese   2400.000000      1997.0         8.4
3005                  45.0  Hungarian   2500.000000      2005.0         7.1
3859                 131.0     Korean   4200.000000      2005.0         7.7
2988                 279.0     Korean  12215.500000      2006.0         7.0

      movie_facebook_likes        profit    genre_1    genre_2
0                    33000    523.505847     Action  Adventure
29                  150000    502.177271     Action  Adventure
26                   26000    458.672302      Drama    Romance
3024                 33000    449.935665     Action  Adventure
3080                 34000    424.449459     Family     Sci-Fi
…                      …           …          …          …
2334                   973  -2127.109510     Action  Adventure
2323                 11000  -2397.701809  Adventure  Animation
3005                   607  -2499.804112      Drama    Romance
3859                  4000  -4199.788333      Crime      Drama
2988                  7000 -12213.298588     Comedy      Drama

[3856 rows x 16 columns]
```

[38]:
```
# Group the dataframe using 'genre_1' as the primary column and 'genre_2' as␣
↪secondary

movies_by_segment = movies.groupby(['genre_1', 'genre_2'])
```

[39]:
```
# Create a new dataframe PopGenre which contains the 'mean' of the gross values␣
↪of each combination of genres present
# Sort this dataframe using the 'gross' column and use index-based positioning␣
↪to find out the five most popular genre combos

PopGenre = pd.DataFrame(movies_by_segment['gross'].mean()).sort_values(by =␣
↪'gross', ascending = False)
PopGenre.iloc[:5, ]
```

[39]:

genre_1	genre_2	gross
Family	Sci-Fi	434.949459
Adventure	Sci-Fi	228.627758
	Family	118.919540
	Animation	116.998550
Action	Adventure	109.595465

Subtask 3.7: Find the critic-favorite and audience-favorite actors

1. Create three new dataframes namely, Meryl_Streep, Leo_Caprio, and Brad_Pitt which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors. Use only the actor_1_name column for extraction. Also, make sure that you use the names 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' for the said extraction.
2. Append the rows of all these dataframes and store them in a new dataframe named Combined.
3. Group the combined dataframe using the actor_1_name column.
4. Find the mean of the num_critic_for_reviews and num_user_for_review and identify the actors which have the highest mean.

[43]:
```
# Create a new dataframe containing Meryl Streep movies in which she is the␣
↪lead actor

Meryl_Streep = movies.loc[movies.actor_1_name == 'Meryl Streep']
Meryl_Streep
```

[43]:

	director_name	num_critic_for_reviews	gross	\
1408	David Frankel	208.0	124.732962	
1575	Sydney Pollack	66.0	87.100000	
1204	Nora Ephron	252.0	94.125426	
1618	David Frankel	234.0	63.536011	
410	Nancy Meyers	187.0	112.703470	
2781	Phyllida Lloyd	331.0	29.959436	
1925	Stephen Daldry	174.0	41.597830	

```
3135   Robert Altman                   211.0   20.338609
1106   Curtis Hanson                    42.0   46.815748
1674   Carl Franklin                    64.0   23.209440
1483  Robert Redford                   227.0   14.998070

                                  genres  actor_1_name  \
1408             [Comedy, Drama, Romance]  Meryl Streep
1575          [Biography, Drama, Romance]  Meryl Streep
1204          [Biography, Drama, Romance]  Meryl Streep
1618             [Comedy, Drama, Romance]  Meryl Streep
410              [Comedy, Drama, Romance]  Meryl Streep
2781          [Biography, Drama, History]  Meryl Streep
1925                    [Drama, Romance]  Meryl Streep
3135               [Comedy, Drama, Music]  Meryl Streep
1106  [Action, Adventure, Crime, Thriller]  Meryl Streep
1674                             [Drama]  Meryl Streep
1483               [Drama, Thriller, War]  Meryl Streep

                  movie_title  num_voted_users  num_user_for_reviews  \
1408    The Devil Wears Prada           286178                 631.0
1575            Out of Africa            52339                 200.0
1204            Julie & Julia            79264                 277.0
1618             Hope Springs            34258                 178.0
410          It's Complicated            69860                 214.0
2781            The Iron Lady            82327                 350.0
1925                The Hours           102123                 660.0
3135  A Prairie Home Companion            19655                 280.0
1106           The River Wild            32544                  69.0
1674           One True Thing             9283                 112.0
1483          Lions for Lambs            41170                 298.0

     language  budget  title_year  imdb_score  movie_facebook_likes  \
1408  English    35.0      2006.0         6.8                     0
1575  English    31.0      1985.0         7.2                     0
1204  English    40.0      2009.0         7.0                 13000
1618  English    30.0      2012.0         6.3                     0
410   English    85.0      2009.0         6.6                     0
2781  English    13.0      2011.0         6.4                 18000
1925  English    25.0      2002.0         7.6                     0
3135  English    10.0      2006.0         6.8                   683
1106  English    45.0      1994.0         6.3                     0
1674  English    30.0      1998.0         7.0                   592
1483  English    35.0      2007.0         6.2                     0

        profit    genre_1    genre_2
1408  89.732962     Comedy      Drama
1575  56.100000  Biography      Drama
```

```
1204  54.125426  Biography      Drama
1618  33.536011     Comedy      Drama
410   27.703470     Comedy      Drama
2781  16.959436  Biography      Drama
1925  16.597830      Drama    Romance
3135  10.338609     Comedy      Drama
1106   1.815748     Action  Adventure
1674  -6.790560      Drama      Drama
1483 -20.001930      Drama   Thriller
```

[45]:
```
# Create a new dataframe containing Leonardo DiCaprio movies in which he is the lead actor

Leonardo_DiCaprio = movies.loc[movies.actor_1_name == 'Leonardo DiCaprio']
Leonardo_DiCaprio
```

[45]:
```
              director_name  num_critic_for_reviews       gross  \
26             James Cameron                   315.0  658.672302
97         Christopher Nolan                   642.0  292.568851
911         Steven Spielberg                   194.0  164.435221
296        Quentin Tarantino                   765.0  162.804648
179   Alejandro G. Iñárritu                   556.0  183.635922
452          Martin Scorsese                   490.0  127.968405
361          Martin Scorsese                   352.0  132.373442
50              Baz Luhrmann                   490.0  144.812796
3476            Baz Luhrmann                   490.0  144.812796
2757            Baz Luhrmann                   106.0   46.338728
1422         Randall Wallace                    83.0   56.876365
308          Martin Scorsese                   606.0  116.866727
1453          Clint Eastwood                   392.0   37.304950
257          Martin Scorsese                   267.0  102.608827
2067              Jerry Zaks                    45.0   12.782508
990              Danny Boyle                   118.0   39.778599
1114              Sam Mendes                   323.0   22.877808
1560               Sam Raimi                    63.0   18.636537
326          Martin Scorsese                   233.0   77.679638
641             Ridley Scott                   238.0   39.380442
307             Edward Zwick                   166.0   57.366262

                                      genres       actor_1_name  \
26                          [Drama, Romance]  Leonardo DiCaprio
97    [Action, Adventure, Sci-Fi, Thriller]  Leonardo DiCaprio
911               [Biography, Crime, Drama]  Leonardo DiCaprio
296                         [Drama, Western]  Leonardo DiCaprio
179   [Adventure, Drama, Thriller, Western]  Leonardo DiCaprio
452                      [Mystery, Thriller]  Leonardo DiCaprio
361                [Crime, Drama, Thriller]  Leonardo DiCaprio
```

```
50                        [Drama, Romance]  Leonardo DiCaprio
3476                      [Drama, Romance]  Leonardo DiCaprio
2757                      [Drama, Romance]  Leonardo DiCaprio
1422                   [Action, Adventure]  Leonardo DiCaprio
308       [Biography, Comedy, Crime, Drama]  Leonardo DiCaprio
1453              [Biography, Crime, Drama]  Leonardo DiCaprio
257                     [Biography, Drama]  Leonardo DiCaprio
2067                                [Drama]  Leonardo DiCaprio
990             [Adventure, Drama, Thriller]  Leonardo DiCaprio
1114                      [Drama, Romance]  Leonardo DiCaprio
1560            [Action, Thriller, Western]  Leonardo DiCaprio
326                         [Crime, Drama]  Leonardo DiCaprio
641               [Action, Drama, Thriller]  Leonardo DiCaprio
307             [Adventure, Drama, Thriller]  Leonardo DiCaprio

                    movie_title  num_voted_users  num_user_for_reviews  \
26                     Titanic           793059                2528.0
97                   Inception          1468200                2803.0
911         Catch Me If You Can           525801                 667.0
296            Django Unchained           955174                1193.0
179                The Revenant           406020                1188.0
452              Shutter Island           786092                 964.0
361                The Departed           873649                2054.0
50              The Great Gatsby           362912                 753.0
3476            The Great Gatsby           362933                 753.0
2757              Romeo + Juliet           167750                 506.0
1422  The Man in the Iron Mask           125219                 244.0
308    The Wolf of Wall Street           780588                1138.0
1453                   J. Edgar           102728                 279.0
257                 The Aviator           264318                 799.0
2067              Marvin's Room            20163                  71.0
990                   The Beach           176169                 548.0
1114         Revolutionary Road           152591                 414.0
1560    The Quick and the Dead            69197                 216.0
326           Gangs of New York           314033                1166.0
641                Body of Lies           174248                 263.0
307                Blood Diamond           400292                 657.0

     language  budget  title_year  imdb_score  movie_facebook_likes  \
26    English   200.0      1997.0         7.7                 26000
97    English   160.0      2010.0         8.8                175000
911   English    52.0      2002.0         8.0                 15000
296   English   100.0      2012.0         8.5                199000
179   English   135.0      2015.0         8.1                190000
452   English    80.0      2010.0         8.1                 53000
361   English    90.0      2006.0         8.5                 29000
50    English   105.0      2013.0         7.3                115000
```

```
3476  English   105.0      2013.0         7.3                115000
2757  English    14.5      1996.0         6.8                 10000
1422  English    35.0      1998.0         6.4                     0
308   English   100.0      2013.0         8.2                138000
1453  English    35.0      2011.0         6.6                 16000
257   English   110.0      2004.0         7.5                     0
2067  English    23.0      1996.0         6.7                  1000
990   English    50.0      2000.0         6.6                     0
1114  English    35.0      2008.0         7.3                     0
1560  English    32.0      1995.0         6.4                     0
326   English   100.0      2002.0         7.5                     0
641   English    70.0      2008.0         7.1                     0
307   English   100.0      2006.0         8.0                 14000

          profit    genre_1    genre_2
26    458.672302      Drama    Romance
97    132.568851     Action  Adventure
911   112.435221  Biography      Crime
296    62.804648      Drama    Western
179    48.635922  Adventure      Drama
452    47.968405    Mystery   Thriller
361    42.373442      Crime      Drama
50     39.812796      Drama    Romance
3476   39.812796      Drama    Romance
2757   31.838728      Drama    Romance
1422   21.876365     Action  Adventure
308    16.866727  Biography     Comedy
1453    2.304950  Biography      Crime
257    -7.391173  Biography      Drama
2067  -10.217492      Drama      Drama
990   -10.221401  Adventure      Drama
1114  -12.122192      Drama    Romance
1560  -13.363463     Action   Thriller
326   -22.320362      Crime      Drama
641   -30.619558     Action      Drama
307   -42.633738  Adventure      Drama
```

[46]:
```
# Create a new dataframe containing Brad Pitt movies in which he is the lead actor

Brad_Pitt = movies.loc[movies.actor_1_name== 'Brad Pitt']
Brad_Pitt
```

[46]:
```
              director_name  num_critic_for_reviews       gross  \
400       Steven Soderbergh                   186.0  183.405771
255              Doug Liman                   233.0  186.336103
940            Neil Jordan                    120.0  105.264608
```

```
470              David Ayer                   406.0   85.707116
254        Steven Soderbergh                  198.0  125.531634
2204  Alejandro G. Iñárritu                   285.0   34.300771
2682         Andrew Dominik                   414.0   14.938570
2898             Tony Scott                   122.0   12.281500
2333    Angelina Jolie Pitt                   131.0    0.531009
1490         Terrence Malick                   584.0   13.303319
101            David Fincher                   362.0  127.490802
683            David Fincher                   315.0   37.023395
1722         Andrew Dominik                    273.0    3.904982
611      Jean-Jacques Annaud                    76.0   37.901509
792          Patrick Gilmore                    98.0   26.288320
147        Wolfgang Petersen                   220.0  133.228348
382              Tony Scott                    142.0    0.026871

                                                 genres actor_1_name  \
400                                  [Crime, Thriller]    Brad Pitt
255          [Action, Comedy, Crime, Romance, Thriller]    Brad Pitt
940                             [Drama, Fantasy, Horror]    Brad Pitt
470                                 [Action, Drama, War]    Brad Pitt
254                                    [Crime, Thriller]    Brad Pitt
2204                                           [Drama]    Brad Pitt
2682                                 [Crime, Thriller]    Brad Pitt
2898          [Action, Crime, Drama, Romance, Thriller]    Brad Pitt
2333                                  [Drama, Romance]    Brad Pitt
1490                                  [Drama, Fantasy]    Brad Pitt
101                             [Drama, Fantasy, Romance]    Brad Pitt
683                                            [Drama]    Brad Pitt
1722        [Biography, Crime, Drama, History, Western]    Brad Pitt
611         [Adventure, Biography, Drama, History, War]    Brad Pitt
792   [Adventure, Animation, Comedy, Drama, Family, …    Brad Pitt
147                                         [Adventure]    Brad Pitt
382                             [Action, Crime, Thriller]    Brad Pitt

                                          movie_title  num_voted_users  \
400                                  Ocean's Eleven            402645
255                                Mr. & Mrs. Smith            348861
940   Interview with the Vampire: The Vampire Chroni…           239752
470                                             Fury            303185
254                                  Ocean's Twelve            284852
2204                                         Babel            243799
2682                        Killing Them Softly               111625
2898                               True Romance               163492
2333                                 By the Sea                7976
1490                              The Tree of Life             136367
101          The Curious Case of Benjamin Button              459346
683                                   Fight Club             1347461
```

```
1722  The Assassination of Jesse James by the Coward…          136104
611                              Seven Years in Tibet            96385
792                    Sinbad: Legend of the Seven Seas          36144
147                                              Troy           381672
382                                         Spy Game            121259

      num_user_for_reviews language  budget  title_year  imdb_score  \
400                  845.0  English    85.0      2001.0         7.8
255                  798.0  English   120.0      2005.0         6.5
940                  406.0  English    60.0      1994.0         7.6
470                  701.0  English    68.0      2014.0         7.6
254                  627.0  English   110.0      2004.0         6.4
2204                 908.0  English    25.0      2006.0         7.5
2682                 369.0  English    15.0      2012.0         6.2
2898                 460.0  English    13.0      1993.0         8.0
2333                  61.0  English    10.0      2015.0         5.3
1490                 975.0  English    32.0      2011.0         6.7
101                  822.0  English   150.0      2008.0         7.8
683                 2968.0  English    63.0      1999.0         8.8
1722                 415.0  English    30.0      2007.0         7.5
611                  119.0  English    70.0      1997.0         7.0
792                   91.0  English    60.0      2003.0         6.7
147                 1694.0  English   175.0      2004.0         7.2
382                  361.0  English    92.0      2001.0         7.0

      movie_facebook_likes     profit    genre_1    genre_2
400                      0  98.405771      Crime   Thriller
255                      0  66.336103     Action     Comedy
940                  11000  45.264608      Drama    Fantasy
470                  82000  17.707116     Action      Drama
254                      0  15.531634      Crime   Thriller
2204                     0   9.300771      Drama      Drama
2682                 20000  -0.061430      Crime   Thriller
2898                 15000  -0.718500     Action      Crime
2333                     0  -9.468991      Drama    Romance
1490                 39000 -18.696681      Drama    Fantasy
101                  23000 -22.509198      Drama    Fantasy
683                  48000 -25.976605      Drama      Drama
1722                     0 -26.095018  Biography      Crime
611                      0 -32.098491  Adventure  Biography
792                    880 -33.711680  Adventure  Animation
147                      0 -41.771652  Adventure  Adventure
382                      0 -91.973129     Action      Crime
```

[47]:
```
# Combine the three dataframes using 'pd.concat()'

combined= pd.concat([Meryl_Streep, Leonardo_DiCaprio, Brad_Pitt])
```

```
combined
```

[47]:

	director_name	num_critic_for_reviews	gross	\
1408	David Frankel	208.0	124.732962	
1575	Sydney Pollack	66.0	87.100000	
1204	Nora Ephron	252.0	94.125426	
1618	David Frankel	234.0	63.536011	
410	Nancy Meyers	187.0	112.703470	
2781	Phyllida Lloyd	331.0	29.959436	
1925	Stephen Daldry	174.0	41.597830	
3135	Robert Altman	211.0	20.338609	
1106	Curtis Hanson	42.0	46.815748	
1674	Carl Franklin	64.0	23.209440	
1483	Robert Redford	227.0	14.998070	
26	James Cameron	315.0	658.672302	
97	Christopher Nolan	642.0	292.568851	
911	Steven Spielberg	194.0	164.435221	
296	Quentin Tarantino	765.0	162.804648	
179	Alejandro G. Iñárritu	556.0	183.635922	
452	Martin Scorsese	490.0	127.968405	
361	Martin Scorsese	352.0	132.373442	
50	Baz Luhrmann	490.0	144.812796	
3476	Baz Luhrmann	490.0	144.812796	
2757	Baz Luhrmann	106.0	46.338728	
1422	Randall Wallace	83.0	56.876365	
308	Martin Scorsese	606.0	116.866727	
1453	Clint Eastwood	392.0	37.304950	
257	Martin Scorsese	267.0	102.608827	
2067	Jerry Zaks	45.0	12.782508	
990	Danny Boyle	118.0	39.778599	
1114	Sam Mendes	323.0	22.877808	
1560	Sam Raimi	63.0	18.636537	
326	Martin Scorsese	233.0	77.679638	
641	Ridley Scott	238.0	39.380442	
307	Edward Zwick	166.0	57.366262	
400	Steven Soderbergh	186.0	183.405771	
255	Doug Liman	233.0	186.336103	
940	Neil Jordan	120.0	105.264608	
470	David Ayer	406.0	85.707116	
254	Steven Soderbergh	198.0	125.531634	
2204	Alejandro G. Iñárritu	285.0	34.300771	
2682	Andrew Dominik	414.0	14.938570	
2898	Tony Scott	122.0	12.281500	
2333	Angelina Jolie Pitt	131.0	0.531009	
1490	Terrence Malick	584.0	13.303319	
101	David Fincher	362.0	127.490802	
683	David Fincher	315.0	37.023395	

1722	Andrew Dominik	273.0	3.904982	
611	Jean-Jacques Annaud	76.0	37.901509	
792	Patrick Gilmore	98.0	26.288320	
147	Wolfgang Petersen	220.0	133.228348	
382	Tony Scott	142.0	0.026871	

	genres	actor_1_name	\
1408	[Comedy, Drama, Romance]	Meryl Streep	
1575	[Biography, Drama, Romance]	Meryl Streep	
1204	[Biography, Drama, Romance]	Meryl Streep	
1618	[Comedy, Drama, Romance]	Meryl Streep	
410	[Comedy, Drama, Romance]	Meryl Streep	
2781	[Biography, Drama, History]	Meryl Streep	
1925	[Drama, Romance]	Meryl Streep	
3135	[Comedy, Drama, Music]	Meryl Streep	
1106	[Action, Adventure, Crime, Thriller]	Meryl Streep	
1674	[Drama]	Meryl Streep	
1483	[Drama, Thriller, War]	Meryl Streep	
26	[Drama, Romance]	Leonardo DiCaprio	
97	[Action, Adventure, Sci-Fi, Thriller]	Leonardo DiCaprio	
911	[Biography, Crime, Drama]	Leonardo DiCaprio	
296	[Drama, Western]	Leonardo DiCaprio	
179	[Adventure, Drama, Thriller, Western]	Leonardo DiCaprio	
452	[Mystery, Thriller]	Leonardo DiCaprio	
361	[Crime, Drama, Thriller]	Leonardo DiCaprio	
50	[Drama, Romance]	Leonardo DiCaprio	
3476	[Drama, Romance]	Leonardo DiCaprio	
2757	[Drama, Romance]	Leonardo DiCaprio	
1422	[Action, Adventure]	Leonardo DiCaprio	
308	[Biography, Comedy, Crime, Drama]	Leonardo DiCaprio	
1453	[Biography, Crime, Drama]	Leonardo DiCaprio	
257	[Biography, Drama]	Leonardo DiCaprio	
2067	[Drama]	Leonardo DiCaprio	
990	[Adventure, Drama, Thriller]	Leonardo DiCaprio	
1114	[Drama, Romance]	Leonardo DiCaprio	
1560	[Action, Thriller, Western]	Leonardo DiCaprio	
326	[Crime, Drama]	Leonardo DiCaprio	
641	[Action, Drama, Thriller]	Leonardo DiCaprio	
307	[Adventure, Drama, Thriller]	Leonardo DiCaprio	
400	[Crime, Thriller]	Brad Pitt	
255	[Action, Comedy, Crime, Romance, Thriller]	Brad Pitt	
940	[Drama, Fantasy, Horror]	Brad Pitt	
470	[Action, Drama, War]	Brad Pitt	
254	[Crime, Thriller]	Brad Pitt	
2204	[Drama]	Brad Pitt	
2682	[Crime, Thriller]	Brad Pitt	
2898	[Action, Crime, Drama, Romance, Thriller]	Brad Pitt	

```
2333                                [Drama, Romance]          Brad Pitt
1490                                [Drama, Fantasy]          Brad Pitt
101                          [Drama, Fantasy, Romance]          Brad Pitt
683                                          [Drama]          Brad Pitt
1722       [Biography, Crime, Drama, History, Western]          Brad Pitt
611        [Adventure, Biography, Drama, History, War]          Brad Pitt
792   [Adventure, Animation, Comedy, Drama, Family, …          Brad Pitt
147                                       [Adventure]          Brad Pitt
382                          [Action, Crime, Thriller]          Brad Pitt

                                        movie_title  num_voted_users  \
1408                          The Devil Wears Prada           286178
1575                                  Out of Africa            52339
1204                                  Julie & Julia            79264
1618                                   Hope Springs            34258
410                                It's Complicated            69860
2781                                  The Iron Lady            82327
1925                                      The Hours           102123
3135                       A Prairie Home Companion            19655
1106                                The River Wild            32544
1674                                One True Thing             9283
1483                                Lions for Lambs            41170
26                                          Titanic           793059
97                                        Inception          1468200
911                          Catch Me If You Can            525801
296                             Django Unchained            955174
179                                 The Revenant            406020
452                               Shutter Island            786092
361                                 The Departed            873649
50                                The Great Gatsby            362912
3476                              The Great Gatsby            362933
2757                                Romeo + Juliet            167750
1422                       The Man in the Iron Mask            125219
308                         The Wolf of Wall Street            780588
1453                                      J. Edgar            102728
257                                     The Aviator            264318
2067                                Marvin's Room             20163
990                                     The Beach            176169
1114                         Revolutionary Road            152591
1560                       The Quick and the Dead             69197
326                          Gangs of New York            314033
641                               Body of Lies            174248
307                               Blood Diamond            400292
400                               Ocean's Eleven            402645
255                             Mr. & Mrs. Smith            348861
940   Interview with the Vampire: The Vampire Chroni…          239752
470                                       Fury            303185
```

254	Ocean's Twelve	284852	
2204	Babel	243799	
2682	Killing Them Softly	111625	
2898	True Romance	163492	
2333	By the Sea	7976	
1490	The Tree of Life	136367	
101	The Curious Case of Benjamin Button	459346	
683	Fight Club	1347461	
1722	The Assassination of Jesse James by the Coward…	136104	
611	Seven Years in Tibet	96385	
792	Sinbad: Legend of the Seven Seas	36144	
147	Troy	381672	
382	Spy Game	121259	

	num_user_for_reviews	language	budget	title_year	imdb_score	\
1408	631.0	English	35.0	2006.0	6.8	
1575	200.0	English	31.0	1985.0	7.2	
1204	277.0	English	40.0	2009.0	7.0	
1618	178.0	English	30.0	2012.0	6.3	
410	214.0	English	85.0	2009.0	6.6	
2781	350.0	English	13.0	2011.0	6.4	
1925	660.0	English	25.0	2002.0	7.6	
3135	280.0	English	10.0	2006.0	6.8	
1106	69.0	English	45.0	1994.0	6.3	
1674	112.0	English	30.0	1998.0	7.0	
1483	298.0	English	35.0	2007.0	6.2	
26	2528.0	English	200.0	1997.0	7.7	
97	2803.0	English	160.0	2010.0	8.8	
911	667.0	English	52.0	2002.0	8.0	
296	1193.0	English	100.0	2012.0	8.5	
179	1188.0	English	135.0	2015.0	8.1	
452	964.0	English	80.0	2010.0	8.1	
361	2054.0	English	90.0	2006.0	8.5	
50	753.0	English	105.0	2013.0	7.3	
3476	753.0	English	105.0	2013.0	7.3	
2757	506.0	English	14.5	1996.0	6.8	
1422	244.0	English	35.0	1998.0	6.4	
308	1138.0	English	100.0	2013.0	8.2	
1453	279.0	English	35.0	2011.0	6.6	
257	799.0	English	110.0	2004.0	7.5	
2067	71.0	English	23.0	1996.0	6.7	
990	548.0	English	50.0	2000.0	6.6	
1114	414.0	English	35.0	2008.0	7.3	
1560	216.0	English	32.0	1995.0	6.4	
326	1166.0	English	100.0	2002.0	7.5	
641	263.0	English	70.0	2008.0	7.1	
307	657.0	English	100.0	2006.0	8.0	

400	845.0	English	85.0	2001.0	7.8
255	798.0	English	120.0	2005.0	6.5
940	406.0	English	60.0	1994.0	7.6
470	701.0	English	68.0	2014.0	7.6
254	627.0	English	110.0	2004.0	6.4
2204	908.0	English	25.0	2006.0	7.5
2682	369.0	English	15.0	2012.0	6.2
2898	460.0	English	13.0	1993.0	8.0
2333	61.0	English	10.0	2015.0	5.3
1490	975.0	English	32.0	2011.0	6.7
101	822.0	English	150.0	2008.0	7.8
683	2968.0	English	63.0	1999.0	8.8
1722	415.0	English	30.0	2007.0	7.5
611	119.0	English	70.0	1997.0	7.0
792	91.0	English	60.0	2003.0	6.7
147	1694.0	English	175.0	2004.0	7.2
382	361.0	English	92.0	2001.0	7.0

	movie_facebook_likes	profit	genre_1	genre_2
1408	0	89.732962	Comedy	Drama
1575	0	56.100000	Biography	Drama
1204	13000	54.125426	Biography	Drama
1618	0	33.536011	Comedy	Drama
410	0	27.703470	Comedy	Drama
2781	18000	16.959436	Biography	Drama
1925	0	16.597830	Drama	Romance
3135	683	10.338609	Comedy	Drama
1106	0	1.815748	Action	Adventure
1674	592	-6.790560	Drama	Drama
1483	0	-20.001930	Drama	Thriller
26	26000	458.672302	Drama	Romance
97	175000	132.568851	Action	Adventure
911	15000	112.435221	Biography	Crime
296	199000	62.804648	Drama	Western
179	190000	48.635922	Adventure	Drama
452	53000	47.968405	Mystery	Thriller
361	29000	42.373442	Crime	Drama
50	115000	39.812796	Drama	Romance
3476	115000	39.812796	Drama	Romance
2757	10000	31.838728	Drama	Romance
1422	0	21.876365	Action	Adventure
308	138000	16.866727	Biography	Comedy
1453	16000	2.304950	Biography	Crime
257	0	-7.391173	Biography	Drama
2067	1000	-10.217492	Drama	Drama
990	0	-10.221401	Adventure	Drama
1114	0	-12.122192	Drama	Romance

```
1560                     0  -13.363463     Action   Thriller
326                      0  -22.320362      Crime      Drama
641                      0  -30.619558     Action      Drama
307                  14000  -42.633738  Adventure      Drama
400                      0   98.405771      Crime   Thriller
255                      0   66.336103     Action     Comedy
940                  11000   45.264608      Drama    Fantasy
470                  82000   17.707116     Action      Drama
254                      0   15.531634      Crime   Thriller
2204                     0    9.300771      Drama      Drama
2682                 20000   -0.061430      Crime   Thriller
2898                 15000   -0.718500     Action      Crime
2333                     0   -9.468991      Drama    Romance
1490                 39000  -18.696681      Drama    Fantasy
101                  23000  -22.509198      Drama    Fantasy
683                  48000  -25.976605      Drama      Drama
1722                     0  -26.095018  Biography      Crime
611                      0  -32.098491  Adventure  Biography
792                    880  -33.711680  Adventure  Animation
147                      0  -41.771652  Adventure  Adventure
382                      0  -91.973129     Action      Crime
```

[49]:
```
# Group the dataframe by 'actor_1_name'

Combined_by_segment = combined.groupby('actor_1_name')
```

[51]:
```
# Remember that we had some null values for the column 'num_critic_for_reviews'.
  ↪ Make sure that none of these null values are
# present in the new dataframe - 'Combined' that we have created

combined.isnull().sum()
```

[51]:
```
director_name              0
num_critic_for_reviews     0
gross                      0
genres                     0
actor_1_name               0
movie_title                0
num_voted_users            0
num_user_for_reviews       0
language                   0
budget                     0
title_year                 0
imdb_score                 0
movie_facebook_likes       0
profit                     0
genre_1                    0
```

```
genre_2                    0
dtype: int64
```

[53]:
```
# Find the mean of 'num_user_for_reviews' for each of the actor. Notice,␣
↪Leonardo's is the highest

Combined_by_segment['num_user_for_reviews'].mean()
```

[53]:
```
actor_1_name
Brad Pitt             742.352941
Leonardo DiCaprio     914.476190
Meryl Streep          297.181818
Name: num_user_for_reviews, dtype: float64
```

[54]:
```
# Find the mean of 'num_critic_for_reviews' for each of the actor. In this case␣
↪as well, Leonardo is leading

Combined_by_segment['num_critic_for_reviews'].mean()
```

[54]:
```
actor_1_name
Brad Pitt             245.000000
Leonardo DiCaprio     330.190476
Meryl Streep          181.454545
Name: num_critic_for_reviews, dtype: float64
```

[ ]:
```

```