

Assignment-4

Group Members

206-Snehal Bankar

207-Ajay Bhapkar

214-Sayali Deshmukh

```
import pandas as pd
```

```
df=pd.read_csv("/content/MOVIES DATASET1.csv")
```

```
print(df)
```

```
   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
...
5038 Color    Scott Smith                 1.0    87.0
5039 Color         NaN                 43.0    43.0
5040 Color Benjamin Roberds                13.0    76.0
5041 Color   Daniel Hsia                 14.0   100.0
5042 Color    Jon Gunn                 43.0    90.0
```

```
   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
...
5038                2.0            318.0  Daphne Zuniga
5039                NaN            319.0  Valorie Curry
5040                0.0                0.0  Maxwell Moody
5041                0.0            489.0  Daniel Henney
5042             16.0             16.0  Brian Herzlinger
```

```
   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
...
5038           637.0         NaN  Comedy|Drama
5039           841.0         NaN  Crime|Drama|Mystery|Thriller
5040                0.0         NaN  Drama|Horror|Thriller
5041           946.0  10443.0  Comedy|Drama|Romance
5042            86.0  85222.0  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
...
5038  ...           6.0  English  Canada  NaN     NaN
5039  ...          359.0  English  USA  TV-14  NaN
5040  ...           3.0  English  USA  NaN  1400.0
5041  ...           9.0  English  USA  PG-13  NaN
5042  ...          84.0  English  USA  PG  1100.0
```

```
   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
...
5038  2013.0            470.0    7.7     NaN
5039     NaN            593.0    7.5   16.00
5040  2013.0             0.0    6.3     NaN
```

5041	2012.0	719.0	6.3	2.35
5042	2004.0	23.0	6.6	1.85

```

movie_facebook_likes
0      33000
1         0
2     85000
3    164000
4         0
...      ...
5038      84
5039    32000
5040      16
5041     660
5042     456

```

[5043 rows x 28 columns]

#1 Calculate the highest director facebook likes

```
df1=pd.read_csv("/content/MOVIES DATASET1.csv",usecols=['director_facebook_likes'])
```

```

print(df1.max())
director_facebook_likes
0         0.0
1        563.0
2         0.0
3       22000.0
4        131.0
...      ...
5038         2.0
5039        NaN
5040         0.0
5041         0.0
5042        16.0

```

[5043 rows x 1 columns]
director_facebook_likes 23000.0
dtype: float64

#2 Describe budget

```
df1=pd.read_csv("/content/MOVIES DATASET1.csv",usecols=['budget'])
```

```
print(df1.describe())
```

```

budget
count  4.551000e+03
mean   3.975262e+07
std    2.061149e+08
min    2.180000e+02
25%    6.000000e+06
50%    2.000000e+07
75%    4.500000e+07
max    1.221550e+10

```

#3 find the mean of number of people voted

```
df1=pd.read_csv("/content/MOVIES DATASET1.csv",usecols=['num_voted_users'])
```

```
print(df1.mean())
```

num_voted_users 83668.160817
dtype: float64

#4 Calculate median of number of faces in poster

```
df1=pd.read_csv("/content/MOVIES DATASET1.csv",usecols=['facenumber_in_poster'])
```

```
print(df1.median())
```

facenumber_in_poster 1.0
dtype: float64

#5 give minimum number of user for reviews

```
df1=pd.read_csv("/content/MOVIES DATASET1.csv",usecols=['num_user_for_reviews'])
print(df1.min())
```

```
num_user_for_reviews    1.0
dtype: float64
```

#6 find number of movies released in USA in the year 2000

```
df2 = df.groupby(['country','title_year']).get_group(('USA',2000)).count()
print(df2['movie_title'])
```

136

#7 Give duplicates of data

```
df.drop_duplicates()
```

color	director_name	num_critic_reviews	duration	director_facebook_likes	actor3_facebook_likes	actor2_name	actor1_facebook_likes	gross	genres	...	num_user_for_reviews	language	country	content_rating	budget	title_year	actor2_facebook_likes	imdb_score	aspect_ratio	movie_facebook_likes
0	Color	James Cameron	7230	178.0	0.0	855.0	Joel David Moore	10000847.0	Action Adventure Fantasy Sci-Fi	...	30540	English	USA	PG-13	27000000	2009.0	2003.0	7.9	1.78	330000
1	Color	Gore Verbinski	3020	169.0	563.0	1000.0	Orlando Bloom	4000012.0	Action Adventure Fantasy	...	12380	English	USA	PG-13	30000000	2007.0	2000.0	7.1	2.35	0
2	Color	Sam Mendes	6020	148.0	0.0	161.0	Rory Kinnear	10000175.0	Action Adventure Thriller	...	9940	English	UK	PG-13	50000000	2015.0	2003.0	6.8	2.35	85000
3	Color	Christopher Nolan	8130	164.0	22000.0	2300.0	Christian Bale	270000642.0	Action Thriller	...	270110	English	USA	PG-13	50000000	2010.0	2003.0	8.5	2.35	164000

color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross	genres	...	num_user_for_reviews	language	country	content_rating	budget	title_year	actor_2_facebook_likes	imdb_score	aspect_ratio	movie_facebook_likes	
4	NaN	Doug Walker	NaN	NaN	131.0	NaN	Rob Walker	131.0	NaN	Documentary	...	NaN	NaN	NaN	NaN	NaN	NaN	12.0	7.1	NaN	0
...	
5038	Color	Scott Smith	1.0	87.0	2.0	318.0	Daphne Zuniga	637.0	NaN	Comedy Drama	...	6.0	English	Canada	NaN	NaN	2013.0	47.0	7.7	NaN	84
5039	Color	NaN	43.0	43.0	NaN	319.0	Valerie Curry	841.0	NaN	Crime Drama Mystery Thriller	...	359.0	English	USA	TV-14	NaN	NaN	593.0	7.5	16.00	32000
5040	Color	Benjamin Robards	13.0	76.0	0.0	0.0	Maxwell Moody	0.0	NaN	Drama Horror Thriller	...	3.0	English	USA	NaN	1400.0	2013.0	0.0	6.3	NaN	16
5041	Color	Daniel Hsia	14.0	100.0	0.0	489.0	Daniel Henney	946.0	1043.0	Comedy Drama Romance	...	9.0	English	USA	PG-13	NaN	2012.0	71.9	6.3	235	660
5042	Color	Jon Gunn	43.0	90.0	16.0	16.0	Brian Herzlinger	862.0	852.0	Documentary	...	84.0	English	USA	PG	1100.0	2004.0	23.0	6.6	1.85	456

4998 rows × 28 columns

```
#8 show(drop) the duplicated rows
print(df.drop_duplicates())

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
5    Color   Andrew Stanton          462.0    132.0
...    ...           ...           ...
5026 Color   Olivier Assayas           81.0    110.0
5027 Color    Jafar Panahi           64.0     90.0
5033 Color   Shane Carruth          143.0     77.0
5035 Color  Robert Rodriguez           56.0     81.0
5042 Color     Jon Gunn           43.0     90.0

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
5	475.0	530.0	Samantha Morton
...
5026	107.0	45.0	Béatrice Dalle
5027	397.0	0.0	Nargess Mamizadeh
5033	291.0	8.0	David Sullivan
5035	0.0	6.0	Peter Marquardt
5042	16.0	16.0	Brian Herzlinger

	actor_1_facebook_likes	gross \
0	1000.0	760505847.0
1	40000.0	309404152.0
2	11000.0	200074175.0
3	27000.0	448130642.0
5	640.0	73058679.0
...
5026	576.0	136007.0
5027	5.0	673780.0
5033	291.0	424760.0
5035	121.0	2040920.0
5042	86.0	85222.0

	genres ... num_user_for_reviews language \
0	Action Adventure Fantasy Sci-Fi ... 3054.0 English
1	Action Adventure Fantasy ... 1238.0 English
2	Action Adventure Thriller ... 994.0 English
3	Action Thriller ... 2701.0 English
5	Action Adventure Sci-Fi ... 738.0 English
...
5026	Drama Music Romance ... 39.0 French
5027	Drama ... 26.0 Persian
5033	Drama Sci-Fi Thriller ... 371.0 English
5035	Action Crime Drama Romance Thriller ... 130.0 Spanish
5042	Documentary ... 84.0 English

	country	content_rating	budget	title_year	actor_2_facebook_likes \
0	USA	PG-13	237000000.0	2009.0	936.0
1	USA	PG-13	300000000.0	2007.0	5000.0
2	UK	PG-13	245000000.0	2015.0	393.0
3	USA	PG-13	250000000.0	2012.0	23000.0
5	USA	PG-13	263700000.0	2012.0	632.0
...
5026	France	R	4500.0	2004.0	133.0
5027	Iran	Not Rated	10000.0	2000.0	0.0
5033	USA	PG-13	7000.0	2004.0	45.0
5035	USA	R	7000.0	1992.0	20.0
5042	USA	PG	1100.0	2004.0	23.0

	imdb_score	aspect_ratio	movie_facebook_likes
0	7.9	1.78	33000
1	7.1	2.35	0
2	6.8	2.35	85000
3	8.5	2.35	164000
5	6.6	2.35	24000
...
5026	6.9	2.35	171
5027	7.5	1.85	697
5033	7.0	1.85	19000
5035	6.9	1.37	0
5042	6.6	1.85	456

[3723 rows x 28 columns]

#9 give correlation of data

print(df.corr())

	num_critic_for_reviews	duration \
num_critic_for_reviews	1.000000	0.258486
duration	0.258486	1.000000
director_facebook_likes	0.180674	0.173296
actor_3_facebook_likes	0.271646	0.123558
actor_1_facebook_likes	0.190016	0.088449
gross	0.480601	0.250298
num_voted_users	0.624943	0.314765
cast_total_facebook_likes	0.263203	0.123074

facenumber_in_poster	-0.033897	0.013469
num_user_for_reviews	0.609387	0.328403
budget	0.119994	0.074276
title_year	0.275707	-0.135038
actor_2_facebook_likes	0.282306	0.131673
imdb_score	0.305303	0.261662
aspect_ratio	-0.049786	-0.090071
movie_facebook_likes	0.683176	0.196605

director_facebook_likes actor_3_facebook_likes \		
num_critic_for_reviews	0.180674	0.271646
duration	0.173296	0.123558
director_facebook_likes	1.000000	0.120199
actor_3_facebook_likes	0.120199	1.000000
actor_1_facebook_likes	0.090723	0.249927
gross	0.144945	0.308026
num_voted_users	0.297057	0.287239
cast_total_facebook_likes	0.119549	0.473920
facenumber_in_poster	-0.041268	0.099368
num_user_for_reviews	0.221890	0.230189
budget	0.021090	0.047451
title_year	-0.063820	0.096137
actor_2_facebook_likes	0.119601	0.559662
imdb_score	0.170802	0.052633
aspect_ratio	0.001642	-0.003366
movie_facebook_likes	0.162048	0.278844

actor_1_facebook_likes gross num_voted_users \			
num_critic_for_reviews	0.190016	0.480601	0.624943
duration	0.088449	0.250298	0.314765
director_facebook_likes	0.090723	0.144945	0.297057
actor_3_facebook_likes	0.249927	0.308026	0.287239
actor_1_facebook_likes	1.000000	0.154468	0.192804
gross	0.154468	1.000000	0.637271
num_voted_users	0.192804	0.637271	1.000000
cast_total_facebook_likes	0.951661	0.247400	0.265911
facenumber_in_poster	0.072257	-0.027755	-0.026998
num_user_for_reviews	0.145461	0.559958	0.798406
budget	0.022639	0.102179	0.079621
title_year	0.086873	0.030886	0.007397
actor_2_facebook_likes	0.390487	0.262768	0.270790
imdb_score	0.076099	0.198021	0.410965
aspect_ratio	-0.020049	0.069346	-0.014761
movie_facebook_likes	0.135348	0.378082	0.537924

cast_total_facebook_likes facenumber_in_poster \		
num_critic_for_reviews	0.263203	-0.033897
duration	0.123074	0.013469
director_facebook_likes	0.119549	-0.041268
actor_3_facebook_likes	0.473920	0.099368
actor_1_facebook_likes	0.951661	0.072257
gross	0.247400	-0.027755
num_voted_users	0.265911	-0.026998
cast_total_facebook_likes	1.000000	0.091475
facenumber_in_poster	0.091475	1.000000
num_user_for_reviews	0.206923	-0.069018
budget	0.036557	-0.019559
title_year	0.109971	0.061504
actor_2_facebook_likes	0.628404	0.071228
imdb_score	0.085787	-0.062958
aspect_ratio	-0.017885	0.013713
movie_facebook_likes	0.209786	0.008918

num_user_for_reviews budget title_year \			
num_critic_for_reviews	0.609387	0.119994	0.275707
duration	0.328403	0.074276	-0.135038
director_facebook_likes	0.221890	0.021090	-0.063820
actor_3_facebook_likes	0.230189	0.047451	0.096137
actor_1_facebook_likes	0.145461	0.022639	0.086873
gross	0.559958	0.102179	0.030886
num_voted_users	0.798406	0.079621	0.007397
cast_total_facebook_likes	0.206923	0.036557	0.109971
facenumber_in_poster	-0.069018	-0.019559	0.061504
num_user_for_reviews	1.000000	0.084292	-0.003147
budget	0.084292	1.000000	0.045726
title_year	-0.003147	0.045726	1.000000
actor_2_facebook_likes	0.219496	0.044236	0.101890
imdb_score	0.292475	0.030688	-0.209167
aspect_ratio	-0.024719	0.006598	0.159973
movie_facebook_likes	0.400594	0.062039	0.218678

	actor_2_facebook_likes	imdb_score	aspect_ratio \
num_critic_for_reviews	0.282306	0.305303	-0.049786
duration	0.131673	0.261662	-0.090071
director_facebook_likes	0.119601	0.170802	0.001642
actor_3_facebook_likes	0.559662	0.052633	-0.003366
actor_1_facebook_likes	0.390487	0.076099	-0.020049
gross	0.262768	0.198021	0.069346
num_voted_users	0.270790	0.410965	-0.014761
cast_total_facebook_likes	0.628404	0.085787	-0.017885
facenumber_in_poster	0.071228	-0.062958	0.013713
num_user_for_reviews	0.219496	0.292475	-0.024719
budget	0.044236	0.030688	0.006598
title_year	0.101890	-0.209167	0.159973
actor_2_facebook_likes	1.000000	0.083808	-0.007783
imdb_score	0.083808	1.000000	0.059445
aspect_ratio	-0.007783	0.059445	1.000000
movie_facebook_likes	0.243487	0.247049	0.025737

	movie_facebook_likes
num_critic_for_reviews	0.683176
duration	0.196605
director_facebook_likes	0.162048
actor_3_facebook_likes	0.278844
actor_1_facebook_likes	0.135348
gross	0.378082
num_voted_users	0.537924
cast_total_facebook_likes	0.209786
facenumber_in_poster	0.008918
num_user_for_reviews	0.400594
budget	0.062039
title_year	0.218678
actor_2_facebook_likes	0.243487
imdb_score	0.247049
aspect_ratio	0.025737
movie_facebook_likes	1.000000

<ipython-input-8-2a5964d788f6>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

```
print(df.corr())
```

#10

```
print(df.cov())
```

	duration \
num_critic_for_reviews	1.478697e+04 7.864841e+02
duration	7.864841e+02 6.349110e+02
director_facebook_likes	6.209067e+04 1.102137e+04
actor_3_facebook_likes	5.521384e+04 5.178387e+03
actor_1_facebook_likes	3.484462e+05 3.352256e+04
gross	4.033268e+09 3.888758e+08
num_voted_users	1.055737e+07 1.099416e+06
cast_total_facebook_likes	5.835781e+05 5.639393e+04
facenumber_in_poster	-8.285428e+00 6.798476e-01
num_user_for_reviews	2.805022e+04 3.115552e+03
budget	3.057919e+09 3.519786e+08
title_year	4.189583e+02 -3.799545e+01
actor_2_facebook_likes	1.393195e+05 1.340403e+04
imdb_score	4.163596e+01 7.407159e+00
aspect_ratio	-8.286055e+00 -3.094366e+00
movie_facebook_likes	1.611872e+06 9.583554e+04

	director_facebook_likes	actor_3_facebook_likes \
num_critic_for_reviews	6.209067e+04	5.521384e+04
duration	1.102137e+04	5.178387e+03
director_facebook_likes	7.914818e+06	5.695669e+05
actor_3_facebook_likes	5.695669e+05	2.772364e+06
actor_1_facebook_likes	3.869475e+06	6.258870e+06
gross	2.923878e+10	3.821700e+10
num_voted_users	1.166379e+08	6.633835e+07
cast_total_facebook_likes	6.160553e+06	1.435699e+07
facenumber_in_poster	-2.339780e+02	3.340384e+02
num_user_for_reviews	2.380619e+05	1.452683e+05
budget	1.264972e+10	1.703143e+10
title_year	-2.240632e+03	2.012353e+03
actor_2_facebook_likes	1.374327e+06	3.770121e+06
imdb_score	5.359587e+02	9.855354e+01
aspect_ratio	3.741410e+00	-7.924615e+00
movie_facebook_likes	8.865869e+06	8.987813e+06

	actor_1_facebook_likes	gross \
num_critic_for_reviews	3.484462e+05	4.033268e+09
duration	3.352256e+04	3.888758e+08
director_facebook_likes	3.869475e+06	2.923878e+10
actor_3_facebook_likes	6.258870e+06	3.821700e+10
actor_1_facebook_likes	2.256232e+08	1.591798e+11
gross	1.591798e+11	4.685812e+15
num_voted_users	4.012392e+08	6.423952e+12
cast_total_facebook_likes	2.597744e+08	3.135947e+11
facenumber_in_poster	2.188860e+03	-3.873446e+06
num_user_for_reviews	8.274949e+05	1.531271e+10
budget	7.278156e+10	1.586166e+15
title_year	1.643128e+04	2.103090e+07
actor_2_facebook_likes	2.372204e+07	7.864260e+10
imdb_score	1.286045e+03	1.426249e+07
aspect_ratio	-4.017976e+02	1.671436e+06
movie_facebook_likes	3.930285e+07	5.380181e+11

	num_voted_users	cast_total_facebook_likes \
num_critic_for_reviews	1.055737e+07	5.835781e+05
duration	1.099416e+06	5.639393e+04
director_facebook_likes	1.166379e+08	6.160553e+06
actor_3_facebook_likes	6.633835e+07	1.435699e+07
actor_1_facebook_likes	4.012392e+08	2.597744e+08
gross	6.423952e+12	3.135947e+11
num_voted_users	1.917817e+10	6.688774e+08
cast_total_facebook_likes	6.688774e+08	3.299236e+08
facenumber_in_poster	-7.523671e+03	3.349224e+03
num_user_for_reviews	4.184799e+07	1.422963e+06
budget	2.358584e+12	1.419869e+11
title_year	1.288219e+04	2.513564e+04
actor_2_facebook_likes	1.517246e+08	4.618384e+07
imdb_score	6.403329e+04	1.753183e+03
aspect_ratio	-2.893018e+03	-4.406113e+02
movie_facebook_likes	1.439267e+09	7.362066e+07

	facenumber_in_poster	num_user_for_reviews \
num_critic_for_reviews	-8.285428e+00	2.805022e+04
duration	6.798476e-01	3.115552e+03
director_facebook_likes	-2.339780e+02	2.380619e+05
actor_3_facebook_likes	3.340384e+02	1.452683e+05
actor_1_facebook_likes	2.188860e+03	8.274949e+05
gross	-3.873446e+06	1.531271e+10
num_voted_users	-7.523671e+03	4.184799e+07
cast_total_facebook_likes	3.349224e+03	1.422963e+06
facenumber_in_poster	4.054488e+00	-5.256856e+01
num_user_for_reviews	-5.256856e+01	1.428711e+05
budget	-8.184479e+06	6.797551e+09
title_year	1.544939e+00	-1.494898e+01
actor_2_facebook_likes	5.809123e+02	3.361557e+05
imdb_score	-1.425547e-01	1.242147e+02
aspect_ratio	3.764651e-02	-1.316649e+01
movie_facebook_likes	3.449754e+02	2.930653e+06

	budget	title_year	actor_2_facebook_likes \
num_critic_for_reviews	3.057919e+09	4.189583e+02	1.393195e+05
duration	3.519786e+08	-3.799545e+01	1.340403e+04
director_facebook_likes	1.264972e+10	-2.240632e+03	1.374327e+06
actor_3_facebook_likes	1.703143e+10	2.012353e+03	3.770121e+06
actor_1_facebook_likes	7.278156e+10	1.643128e+04	2.372204e+07
gross	1.586166e+15	2.103090e+07	7.864260e+10
num_voted_users	2.358584e+12	1.288219e+04	1.517246e+08
cast_total_facebook_likes	1.419869e+11	2.513564e+04	4.618384e+07
facenumber_in_poster	-8.184479e+06	1.544939e+00	5.809123e+02
num_user_for_reviews	6.797551e+09	-1.494898e+01	3.361557e+05
budget	4.248335e+16	1.176608e+08	3.842339e+10
title_year	1.176608e+08	1.556156e+02	5.189927e+03
actor_2_facebook_likes	3.842339e+10	5.189927e+03	1.634131e+07
imdb_score	7.056363e+06	-2.908298e+00	3.809056e+02
aspect_ratio	1.093616e+06	1.589023e+00	-4.426008e+01
movie_facebook_likes	2.566124e+11	5.306848e+04	1.903769e+07

	imdb_score	aspect_ratio	movie_facebook_likes
num_critic_for_reviews	4.163596e+01	-8.286055e+00	1.611872e+06
duration	7.407159e+00	-3.094366e+00	9.583554e+04
director_facebook_likes	5.359587e+02	3.741410e+00	8.865869e+06
actor_3_facebook_likes	9.855354e+01	-7.924615e+00	8.987813e+06
actor_1_facebook_likes	1.286045e+03	-4.017976e+02	3.930285e+07
gross	1.426249e+07	1.671436e+06	5.380181e+11
num_voted_users	6.403329e+04	-2.893018e+03	1.439267e+09
cast_total_facebook_likes	1.753183e+03	-4.406113e+02	7.362066e+07


```

facenumber_in_poster  -1.425547e-01  3.764651e-02   3.449754e+02
num_user_for_reviews   1.242147e+02 -1.316649e+01   2.930653e+06
budget                 7.056363e+06  1.093616e+06   2.566124e+11
title_year             -2.908298e+00  1.589023e+00   5.306848e+04
actor_2_facebook_likes  3.809056e+02 -4.426008e+01   1.903769e+07
imdb_score             1.265886e+00  9.027006e-02   5.370276e+03
aspect_ratio           9.027006e-02  1.918537e+00   7.088611e+02
movie_facebook_likes    5.370276e+03  7.088611e+02   3.732796e+08

```

```
#11 convert name of actor_2 column in lowercase
```

```
print(df['actor_2_name'].str.lower())
```

```

0    joel david moore
1    orlando bloom
2    rory kinnear
3    christian bale
4    rob walker
...
5038    daphne zuniga
5039    valorie curry
5040    maxwell moody
5041    daniel henney
5042    brian herzlinger
Name: actor_2_name, Length: 5043, dtype: object

```

```
#12 sorting the budget
```

```
sorted_df=df.sort_values(by=["budget"])
```

```
print(sorted_df)
```

```

color    director_name num_critic_for_reviews duration \
4799 Color Jonathan Caouette          72.0    88.0
5042 Color Jon Gunn          43.0    90.0
5040 Color Benjamin Roberds          13.0    76.0
5036 Color Anthony Vallone           NaN    84.0
5026 Color Olivier Assayas          81.0   110.0
... ..
5030 Color Tadeo Garcia           NaN    84.0
5032 Color Ash Baron-Cohen         10.0    98.0
5038 Color Scott Smith            1.0    87.0
5039 Color NaN          43.0    43.0
5041 Color Daniel Hsia           14.0   100.0

director_facebook_likes actor_3_facebook_likes actor_2_name \
4799          20.0          0.0 Jonathan Caouette
5042          16.0          16.0 Brian Herzlinger
5040           0.0           0.0 Maxwell Moody
5036           2.0           2.0 John Considine
5026         107.0          45.0 Béatrice Dalle
... ..
5030           5.0          12.0 Michael Cortez
5032           3.0         152.0 Stanley B. Herman
5038           2.0         318.0 Daphne Zuniga
5039          NaN         319.0 Valorie Curry
5041           0.0         489.0 Daniel Henney

actor_1_facebook_likes gross genres ... \
4799          58.0 592014.0 Biography|Documentary ...
5042          86.0 85222.0 Documentary ...
5040           0.0 NaN Drama|Horror|Thriller ...
5036          45.0 NaN Crime|Drama ...
5026         576.0 136007.0 Drama|Music|Romance ...
... ..
5030          21.0 NaN Drama ...
5032         789.0 NaN Crime|Drama ...
5038         637.0 NaN Comedy|Drama ...
5039         841.0 NaN Crime|Drama|Mystery|Thriller ...
5041         946.0 10443.0 Comedy|Drama|Romance ...

num_user_for_reviews language country content_rating budget \
4799         114.0 English USA Unrated 218.0
5042          84.0 English USA PG 1100.0
5040           3.0 English USA NaN 1400.0
5036           1.0 English USA PG-13 3250.0
5026          39.0 French France R 4500.0
... ..
5030           3.0 English USA NaN NaN

```

5032	14.0	English	USA	NaN	NaN
5038	6.0	English	Canada	NaN	NaN
5039	359.0	English	USA	TV-14	NaN
5041	9.0	English	USA	PG-13	NaN

	title_year	actor_2_facebook_likes	imdb_score	aspect_ratio	\
4799	2003.0	20.0	7.2	1.37	
5042	2004.0	23.0	6.6	1.85	
5040	2013.0	0.0	6.3	NaN	
5036	2005.0	44.0	7.8	NaN	
5026	2004.0	133.0	6.9	2.35	
...	
5030	2004.0	20.0	6.1	NaN	
5032	1995.0	194.0	6.4	NaN	
5038	2013.0	470.0	7.7	NaN	
5039	NaN	593.0	7.5	16.00	
5041	2012.0	719.0	6.3	2.35	

	movie_facebook_likes
4799	754
5042	456
5040	16
5036	4
5026	171
...	...
5030	22
5032	20
5038	84
5039	32000
5041	660

[5043 rows x 28 columns]

#13 counting the number of black and white movies

```
df3 = df.groupby('color').get_group(' Black and White').count()
print(df3['movie_title'])
```

209

#14 find the number of movies released in japanese language in the data

```
a=df.groupby('language').get_group('Japanese')
```

```
print(a['actor_1_name'])
98      Mark Chinnery
204      Mark Chinnery
1517     Rumi Hiiragi
1761     Shun Oguri
2047     Christian Bale
2241     Pablo Sevilla
2323     Minnie Driver
2334     William Hootkins
2373     Bunta Sugawara
2863     Yuki Matsuzaki
3012     Takuya Kimura
3020     Tatsuo Matsumura
3423     Mitsuo Iwata
3684     Lindsay Kay Hayward
4064     Veronica Taylor
4542     Hiroshi Abe
4747     Takashi Shimura
5029     Kôji Yakusho
```

#15 check for missing values in dataframe

```
print(df.isnull().count())
```

color	5043
director_name	5043
num_critic_for_reviews	5043
duration	5043
director_facebook_likes	5043
actor_3_facebook_likes	5043
actor_2_name	5043
actor_1_facebook_likes	5043

```
gross          5043
genres         5043
actor_1_name   5043
movie_title    5043
num_voted_users 5043
cast_total_facebook_likes 5043
actor_3_name   5043
facenumber_in_poster 5043
plot_keywords  5043
movie_imdb_link 5043
num_user_for_reviews 5043
language       5043
country        5043
content_rating 5043
budget         5043
title_year     5043
actor_2_facebook_likes 5043
imdb_score     5043
aspect_ratio   5043
movie_facebook_likes 5043
```

#16 Drop rows with missing values

```
print(df.dropna(inplace=True))
```

None

#17 check for missing values in the data

```
print(df.isna())
color director_name num_critic_for_reviews duration \
0  False      False      False      False
1  False      False      False      False
2  False      False      False      False
3  False      False      False      False
5  False      False      False      False
...  ...      ...      ...      ...
5026 False      False      False      False
5027 False      False      False      False
5033 False      False      False      False
5035 False      False      False      False
5042 False      False      False      False

director_facebook_likes actor_3_facebook_likes actor_2_name \
0      False      False      False
1      False      False      False
2      False      False      False
3      False      False      False
5      False      False      False
...      ...      ...      ...
5026      False      False      False
5027      False      False      False
5033      False      False      False
5035      False      False      False
5042      False      False      False

actor_1_facebook_likes gross genres ... num_user_for_reviews \
0      False False False ...      False
1      False False False ...      False
2      False False False ...      False
3      False False False ...      False
5      False False False ...      False
...      ...  ...  ...      ...
5026      False False False ...      False
5027      False False False ...      False
5033      False False False ...      False
5035      False False False ...      False
5042      False False False ...      False

language country content_rating budget title_year \
0  False False      False False      False
1  False False      False False      False
2  False False      False False      False
3  False False      False False      False
5  False False      False False      False
...  ...  ...      ...  ...      ...
5026 False False      False False      False
5027 False False      False False      False
5033 False False      False False      False
```

5035	False	False	False	False	False
5042	False	False	False	False	False

	actor_2_facebook_likes	imdb_score	aspect_ratio	movie_facebook_likes
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
5	False	False	False	False
...
5026	False	False	False	False
5027	False	False	False	False
5033	False	False	False	False
5035	False	False	False	False
5042	False	False	False	False

[3756 rows x 28 columns]

#18 fill missing values with specific value (0)

```
print(df.fillna(0))
```

	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	
5	Color	Andrew Stanton	462.0	132.0	
...	
5026	Color	Olivier Assayas	81.0	110.0	
5027	Color	Jafar Panahi	64.0	90.0	
5033	Color	Shane Carruth	143.0	77.0	
5035	Color	Robert Rodriguez	56.0	81.0	
5042	Color	Jon Gunn	43.0	90.0	

	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	
5	475.0	530.0	Samantha Morton	
...	
5026	107.0	45.0	Béatrice Dalle	
5027	397.0	0.0	Nargess Mamizadeh	
5033	291.0	8.0	David Sullivan	
5035	0.0	6.0	Peter Marquardt	
5042	16.0	16.0	Brian Herzlinger	

	actor_1_facebook_likes	gross	\
0	1000.0	760505847.0	
1	40000.0	309404152.0	
2	11000.0	200074175.0	
3	27000.0	448130642.0	
5	640.0	73058679.0	
...	
5026	576.0	136007.0	
5027	5.0	673780.0	
5033	291.0	424760.0	
5035	121.0	2040920.0	
5042	86.0	85222.0	

	genres	...	num_user_for_reviews	language	\
0	Action Adventure Fantasy Sci-Fi	...	3054.0	English	
1	Action Adventure Fantasy	...	1238.0	English	
2	Action Adventure Thriller	...	994.0	English	
3	Action Thriller	...	2701.0	English	
5	Action Adventure Sci-Fi	...	738.0	English	
...	
5026	Drama Music Romance	...	39.0	French	
5027	Drama	...	26.0	Persian	
5033	Drama Sci-Fi Thriller	...	371.0	English	
5035	Action Crime Drama Romance Thriller	...	130.0	Spanish	
5042	Documentary	...	84.0	English	

	country	content_rating	budget	title_year	actor_2_facebook_likes	\
0	USA	PG-13	237000000.0	2009.0	936.0	
1	USA	PG-13	300000000.0	2007.0	5000.0	
2	UK	PG-13	245000000.0	2015.0	393.0	
3	USA	PG-13	250000000.0	2012.0	23000.0	
5	USA	PG-13	263700000.0	2012.0	632.0	
...	
5026	France	R	4500.0	2004.0	133.0	

5027	Iran	Not Rated	10000.0	2000.0	0.0
5033	USA	PG-13	7000.0	2004.0	45.0
5035	USA	R	7000.0	1992.0	20.0
5042	USA	PG	1100.0	2004.0	23.0

	imdb_score	aspect_ratio	movie_facebook_likes
0	7.9	1.78	33000
1	7.1	2.35	0
2	6.8	2.35	85000
3	8.5	2.35	164000
5	6.6	2.35	24000
...
5026	6.9	2.35	171
5027	7.5	1.85	697
5033	7.0	1.85	19000
5035	6.9	1.37	0
5042	6.6	1.85	456

[3756 rows x 28 columns]

#19 show the actor_1_facebook_likes which are greater than 100000 (filter rows based on a condition)

```
print(df[df['actor_1_facebook_likes']>100000])
```

color	director_name	num_critic_for_reviews	duration	\
1223	Color	David R. Ellis	221.0	82.0
1902	Color	Adam McKay	181.0	98.0
2269	Color	Nimród Antal	107.0	88.0
2939	Color	John Erick Dowdle	198.0	89.0
4592	Color	Johnny Remo	2.0	112.0

	director_facebook_likes	actor_3_facebook_likes	actor_2_name	\
1223	160.0	748.0	Andrew Fiscella	
1902	285.0	7000.0	Will Ferrell	
2269	190.0	163.0	Fred Ward	
2939	66.0	1000.0	Jay Hernandez	
4592	74.0	891.0	Randy Wayne	

	actor_1_facebook_likes	gross	genres	...	\
1223	164000.0	66466372.0	Horror	...	
1902	640000.0	84136909.0	Comedy	...	
2269	137000.0	15988876.0	Action Crime Thriller	...	
2939	137000.0	31691811.0	Horror Sci-Fi Thriller	...	
4592	260000.0	96734.0	Action Drama	...	

	num_user_for_reviews	language	country	content_rating	budget	\
1223	290.0	English	USA	R	40000000.0	
1902	577.0	English	USA	PG-13	26000000.0	
2269	110.0	English	USA	PG-13	27000000.0	
2939	369.0	English	USA	R	12000000.0	
4592	5.0	English	USA	PG-13	1000000.0	

	title_year	actor_2_facebook_likes	imdb_score	aspect_ratio	\
1223	2009.0	137000.0	5.2	2.35	
1902	2004.0	8000.0	7.2	1.85	
2269	2009.0	459.0	5.7	2.35	
2939	2008.0	1000.0	6.0	1.85	
4592	2012.0	984.0	5.6	1.85	

	movie_facebook_likes
1223	0
1902	0
2269	0
2939	0
4592	706

[5 rows x 28 columns]

#20 count the number of movies realeased in perticular country (count the accurrences of each unique country in a column)

```
print(df['country'].value_counts())
```

USA	2987
UK	318
France	101
Germany	80
Canada	59
Australia	39
Spain	21
Japan	15

```

Hong Kong      13
China          13
Italy          11
New Zealand    11
South Korea    8
Denmark        8
Ireland        7
Mexico         6
Brazil         5
India          5
Iran           4
Thailand       4
Norway         4
Russia         3
Argentina      3
Netherlands   3
South Africa   3
Czech Republic 3
Hungary        2
Taiwan         2
Romania        2
Finland        1
Indonesia      1
Israel         1
Poland         1
Colombia       1
New Line       1
Iceland        1
Aruba          1
Peru           1
Belgium        1
Georgia        1
West Germany   1
Chile          1
Official site  1
Greece         1
Afghanistan    1
Name: country, dtype: int64

```

#21 apply multiple aggregation funtions to each group

```
print(df.groupby('num_voted_users').agg(['mean','max','min']))
```

```

num_critic_for_reviews    duration \
mean max min mean max min
num_voted_users
91      12.0 12.0 12.0  88.0 88.0 88.0
154     29.0 29.0 29.0  93.0 93.0 93.0
241     10.0 10.0 10.0  95.0 95.0 95.0
344     94.0 94.0 94.0 106.0 106.0 106.0
397      7.0  7.0  7.0 120.0 120.0 120.0
...
1324680 215.0 215.0 215.0 178.0 178.0 178.0
1347461 315.0 315.0 315.0 151.0 151.0 151.0
1468200 642.0 642.0 642.0 148.0 148.0 148.0
1676169 645.0 645.0 645.0 152.0 152.0 152.0
1689764 199.0 199.0 199.0 142.0 142.0 142.0

```

```

director_facebook_likes \
mean max min
num_voted_users
91      0.0  0.0  0.0
154      0.0  0.0  0.0
241      0.0  0.0  0.0
344      5.0  5.0  5.0
397      0.0  0.0  0.0
...
1324680 16000.0 16000.0 16000.0
1347461 21000.0 21000.0 21000.0
1468200 22000.0 22000.0 22000.0
1676169 22000.0 22000.0 22000.0
1689764  0.0  0.0  0.0

```

```

actor_3_facebook_likes ... actor_2_facebook_likes imdb_score \
mean ... min mean
num_voted_users
91      3.0 ... 44.0  6.1
154     59.0 ... 169.0  6.8
241    309.0 ... 665.0  4.8
344    356.0 ... 403.0  5.7
397     40.0 ...  69.0  3.3
...

```

1324680	857.0	...	902.0	8.9
1347461	637.0	...	783.0	8.8
1468200	23000.0	...	27000.0	8.8
1676169	11000.0	...	13000.0	9.0
1689764	461.0	...	745.0	9.3

	aspect_ratio			movie_facebook_likes \		
	max	min	mean	max	min	mean
num_voted_users						
91	6.1	6.1	1.78	1.78	1.78	27.0
154	6.8	6.8	16.00	16.00	16.00	287.0
241	4.8	4.8	1.85	1.85	1.85	261.0
344	5.7	5.7	2.35	2.35	2.35	655.0
397	3.3	3.3	1.85	1.85	1.85	32.0
...
1324680	8.9	8.9	2.35	2.35	2.35	45000.0
1347461	8.8	8.8	2.35	2.35	2.35	48000.0
1468200	8.8	8.8	2.35	2.35	2.35	175000.0
1676169	9.0	9.0	2.35	2.35	2.35	37000.0
1689764	9.3	9.3	1.85	1.85	1.85	108000.0

	max	min
num_voted_users		
91	27	27
154	287	287
241	261	261
344	655	655
397	32	32
...
1324680	45000	45000
1347461	48000	48000
1468200	175000	175000
1676169	37000	37000
1689764	108000	108000

[3674 rows x 45 columns]

```
#22 compute the quantile of a facenumber_in_poster
print(df['facenumber_in_poster'].quantile([0.25,0.5,0.75,1]))
```

```
0.25    0.0
0.50    1.0
0.75    2.0
1.00   43.0
```

```
#23 check missing values in data frame
print(df.isnull())
```

	color	director_name	num_critic_for_reviews	duration \
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
5	False	False	False	False
...
5026	False	False	False	False
5027	False	False	False	False
5033	False	False	False	False
5035	False	False	False	False
5042	False	False	False	False

	director_facebook_likes	actor_3_facebook_likes	actor_2_name \
0	False	False	False
1	False	False	False
2	False	False	False
3	False	False	False
5	False	False	False
...
5026	False	False	False
5027	False	False	False
5033	False	False	False
5035	False	False	False
5042	False	False	False

	actor_1_facebook_likes	gross	genres	...	num_user_for_reviews \
0	False	False	False	...	False
1	False	False	False	...	False
2	False	False	False	...	False
3	False	False	False	...	False
5	False	False	False	...	False

```
...
5026      False False False ...      False
5027      False False False ...      False
5033      False False False ...      False
5035      False False False ...      False
5042      False False False ...      False
```

```
language country content_rating budget title_year \
0      False False      False False      False
1      False False      False False      False
2      False False      False False      False
3      False False      False False      False
5      False False      False False      False
...
5026      False False      False False      False
5027      False False      False False      False
5033      False False      False False      False
5035      False False      False False      False
5042      False False      False False      False
```

```
actor_2_facebook_likes imdb_score aspect_ratio movie_facebook_likes
0      False      False      False      False
1      False      False      False      False
2      False      False      False      False
3      False      False      False      False
5      False      False      False      False
...
5026      False      False      False      False
5027      False      False      False      False
5033      False      False      False      False
5035      False      False      False      False
5042      False      False      False      False
```

[3756 rows x 28 columns]

#24 fill null vale with the column mean

```
print(df.fillna(df.mean()))
```

```
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
5      Color      Andrew Stanton      462.0  132.0
...
5026      Color      Olivier Assayas      81.0  110.0
5027      Color      Jafar Panahi      64.0  90.0
5033      Color      Shane Carruth      143.0  77.0
5035      Color      Robert Rodriguez      56.0  81.0
5042      Color      Jon Gunn      43.0  90.0
```

```
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
5      475.0      530.0      Samantha Morton
...
5026      107.0      45.0      Béatrice Dalle
5027      397.0      0.0      Nargess Mamizadeh
5033      291.0      8.0      David Sullivan
5035      0.0      6.0      Peter Marquardt
5042      16.0      16.0      Brian Herzlinger
```

```
actor_1_facebook_likes gross \
0      1000.0  760505847.0
1      40000.0  309404152.0
2      11000.0  200074175.0
3      27000.0  448130642.0
5      640.0  73058679.0
...
5026      576.0  136007.0
5027      5.0  673780.0
5033      291.0  424760.0
5035      121.0  2040920.0
5042      86.0  85222.0
```

```
genres ... num_user_for_reviews language \
0      Action|Adventure|Fantasy|Sci-Fi ...      3054.0      English
1      Action|Adventure|Fantasy ...      1238.0      English
2      Action|Adventure|Thriller ...      994.0      English
```


3		Action Thriller ...		2701.0	English
5		Action Adventure Sci-Fi ...		738.0	English
...	
5026		Drama Music Romance ...		39.0	French
5027		Drama ...		26.0	Persian
5033		Drama Sci-Fi Thriller ...		371.0	English
5035		Action Crime Drama Romance Thriller ...		130.0	Spanish
5042		Documentary ...		84.0	English

	country	content_rating	budget	title_year	actor_2_facebook_likes \
0	USA	PG-13	237000000.0	2009.0	936.0
1	USA	PG-13	300000000.0	2007.0	5000.0
2	UK	PG-13	245000000.0	2015.0	393.0
3	USA	PG-13	250000000.0	2012.0	23000.0
5	USA	PG-13	263700000.0	2012.0	632.0
...
5026	France	R	4500.0	2004.0	133.0
5027	Iran	Not Rated	10000.0	2000.0	0.0
5033	USA	PG-13	7000.0	2004.0	45.0
5035	USA	R	7000.0	1992.0	20.0
5042	USA	PG	1100.0	2004.0	23.0

	imdb_score	aspect_ratio	movie_facebook_likes
0	7.9	1.78	33000
1	7.1	2.35	0
2	6.8	2.35	85000
3	8.5	2.35	164000
5	6.6	2.35	24000
...
5026	6.9	2.35	171
5027	7.5	1.85	697
5033	7.0	1.85	19000
5035	6.9	1.37	0
5042	6.6	1.85	456

#25

```
from scipy import stats
```

```
z_scores=stats.zscore(df['duration'])
```

```
threshold=3
```

```
print(df[(z_scores<threshold)])
```

Empty DataFrame

Columns: [color, director_name, num_critc_for_reviews, duration, director_facebook_likes, actor_3_facebook_likes, actor_2_name, actor_1_facebook_likes, gross, genres, actor_1_name, movie_title, num_voted_users, cast_total_facebook_likes, actor_3_name, facenumber_in_poster, plot_keywords, movie_imdb_link, num_user_for_reviews, language, country, content_rating, budget, title_year, actor_2_facebook_likes, imdb_score, aspect_ratio, movie_facebook_likes]
Index: []

#26

```
print(df.loc[df['language']=='English','language']=='English')
```

```
0    True
1    True
2    True
3    True
5    True
```

...

```
5038    True
5039    True
5040    True
5041    True
5042    True
```

Name: language, Length: 4704, dtype: bool

#27 print all the number of categories in content_rating (print one-hot encoding)

```
print(pd.get_dummies(df['content_rating']))
```

	Approved	G	GP	M	NC-17	Not Rated	PG	PG-13	Passed	R	Unrated	X
0	0	0	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0	0
2	0	0	0	0	0	0	0	1	0	0	0	0
3	0	0	0	0	0	0	0	1	0	0	0	0
5	0	0	0	0	0	0	0	1	0	0	0	0
...
5026	0	0	0	0	0	0	0	0	1	0	0	0
5027	0	0	0	0	0	1	0	0	0	0	0	0
5033	0	0	0	0	0	0	0	1	0	0	0	0
5035	0	0	0	0	0	0	0	0	1	0	0	0

5042 0 0 0 0 0 0 1 0 0 0 0 0

[3756 rows x 12 columns]

CodeText

#28 display the record of first 10 movies

```
print(df.iloc[1:10])
```

	color	director_name	num_critic_for_reviews	duration \
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
5	Color	Andrew Stanton	462.0	132.0
6	Color	Sam Raimi	392.0	156.0
7	Color	Nathan Greno	324.0	100.0
8	Color	Joss Whedon	635.0	141.0
9	Color	David Yates	375.0	153.0

	director_facebook_likes	actor_3_facebook_likes	actor_2_name \
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
5	475.0	530.0	Samantha Morton
6	0.0	4000.0	James Franco
7	15.0	284.0	Donna Murphy
8	0.0	19000.0	Robert Downey Jr.
9	282.0	10000.0	Daniel Radcliffe

	actor_1_facebook_likes	gross \
1	40000.0	309404152.0
2	11000.0	200074175.0
3	27000.0	448130642.0
4	131.0	NaN
5	640.0	73058679.0
6	24000.0	336530303.0
7	799.0	200807262.0
8	26000.0	458991599.0
9	25000.0	301956980.0

	genres ... \
1	Action Adventure Fantasy ...
2	Action Adventure Thriller ...
3	Action Thriller ...
4	Documentary ...
5	Action Adventure Sci-Fi ...
6	Action Adventure Romance ...
7	Adventure Animation Comedy Family Fantasy Musi... ...
8	Action Adventure Sci-Fi ...
9	Adventure Family Fantasy Mystery ...

	num_user_for_reviews	language	country	content_rating	budget \
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
5	738.0	English	USA	PG-13	263700000.0
6	1902.0	English	USA	PG-13	258000000.0
7	387.0	English	USA	PG	260000000.0
8	1117.0	English	USA	PG-13	250000000.0
9	973.0	English	UK	PG	250000000.0

	title_year	actor_2_facebook_likes	imdb_score	aspect_ratio \
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
5	2012.0	632.0	6.6	2.35
6	2007.0	11000.0	6.2	2.35
7	2010.0	553.0	7.8	1.85
8	2015.0	21000.0	7.5	2.35
9	2009.0	11000.0	7.5	2.35

	movie_facebook_likes
1	0
2	85000
3	164000
4	0
5	24000

6	0
7	29000
8	118000
9	10000