

```
In [369... import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

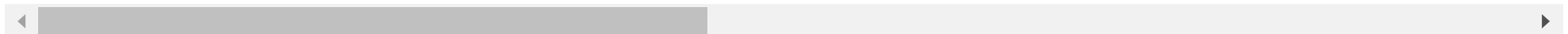
## Importing and reading the movie database. Storing it in a variable called movies.

```
In [370... movies = pd.read_csv('IMDB_Movies.csv')
originaldata = movies
movies
```

Out[370]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	
0	Color	James Cameron	723.0	178.0	0.0	855.0	Joel David Moore	1000.0	7
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0	Orlando Bloom	40000.0	3
2	Color	Sam Mendes	602.0	148.0	0.0	161.0	Rory Kinnear	11000.0	2
3	Color	Christopher Nolan	813.0	164.0	22000.0	23000.0	Christian Bale	27000.0	4
4	NaN	Doug Walker	NaN	NaN	131.0	NaN	Rob Walker	131.0	
...	...	...	...	...	...	...	...	...	
5038	Color	Scott Smith	1.0	87.0	2.0	318.0	Daphne Zuniga	637.0	
5039	Color	NaN	43.0	43.0	NaN	319.0	Valorie Curry	841.0	
5040	Color	Benjamin Roberds	13.0	76.0	0.0	0.0	Maxwell Moody	0.0	
5041	Color	Daniel Hsia	14.0	100.0	0.0	489.0	Daniel Henney	946.0	
5042	Color	Jon Gunn	43.0	90.0	16.0	16.0	Brian Herzlinger	86.0	

5043 rows × 28 columns



## Inspecting the dataframe shapes, columns, variable types etc.

In [371... `movies.shape`

Out[371]: (5043, 28)

In [372... `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                 5023 non-null   object
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(12), int64(3), object(13)
memory usage: 1.1+ MB

```

**Calculating the number of Null values of rows and columns.  
Finding the percentage of Null values in each column.**

```
In [373... movies.isnull().sum(axis=0).sort_values(ascending=False)
```

```
Out[373]: gross            884  
budget            492  
aspect_ratio      329  
content_rating    303  
plot_keywords     153  
title_year        108  
director_name     104  
director_facebook_likes 104  
num_critic_for_reviews 50  
actor_3_name       23  
actor_3_facebook_likes 23  
num_user_for_reviews 20  
color             19  
duration          15  
facenumber_in_poster 13  
actor_2_name       13  
actor_2_facebook_likes 13  
language          12  
actor_1_name        7  
actor_1_facebook_likes 7  
country            5  
cast_total_facebook_likes 0  
num_voted_users    0  
movie_title        0  
movie_imdb_link    0  
genres             0  
imdb_score          0  
movie_facebook_likes 0  
dtype: int64
```

```
In [374... movies.isnull().sum(axis=1).sort_values(ascending=False)
```

```
Out[374]: 279      15
          4       13
          4945    11
          2241    11
          2342    10
          ..
          1703     0
          1702     0
          1701     0
          1700     0
          5042     0
          Length: 5043, dtype: int64
```

```
In [375... movies.isnull().sum(axis=0).sort_values(ascending=False)/len(movies) * 100
```

```

Out[375]: gross                17.529248
          budget                9.756098
          aspect_ratio          6.523895
          content_rating         6.008328
          plot_keywords          3.033908
          title_year             2.141582
          director_name          2.062265
          director_facebook_likes 2.062265
          num_critic_for_reviews 0.991473
          actor_3_name           0.456078
          actor_3_facebook_likes 0.456078
          num_user_for_reviews   0.396589
          color                  0.376760
          duration               0.297442
          facenumber_in_poster   0.257783
          actor_2_name           0.257783
          actor_2_facebook_likes 0.257783
          language               0.237954
          actor_1_name           0.138806
          actor_1_facebook_likes 0.138806
          country                0.099147
          cast_total_facebook_likes 0.000000
          num_voted_users        0.000000
          movie_title            0.000000
          movie_imdb_link        0.000000
          genres                 0.000000
          imdb_score             0.000000
          movie_facebook_likes   0.000000
          dtype: float64

```

## Drop the unnecessary columns.

## Columns which are irrelevant and to be dropped are:

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 movies\_imdb\_link aspect\_ratio plot\_keywords

```
In [376... movies = movies.drop([  
    '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'],axis=1)
```

```
In [377... movies
```

Out[377]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_for_revi
0	James Cameron	723.0	760505847.0	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	3
1	Gore Verbinski	302.0	309404152.0	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End	471220	1
2	Sam Mendes	602.0	200074175.0	Action Adventure Thriller	Christoph Waltz	Spectre	275868	
3	Christopher Nolan	813.0	448130642.0	Action Thriller	Tom Hardy	The Dark Knight Rises	1144337	2
4	Doug Walker	NaN	NaN	Documentary	Doug Walker	Star Wars: Episode VII - The Force Awakens ...	8	
...	...	...	...	...	...	...	...	
5038	Scott Smith	1.0	NaN	Comedy Drama	Eric Mabius	Signed Sealed Delivered	629	
5039	NaN	43.0	NaN	Crime Drama Mystery Thriller	Natalie Zea	The Following	73839	
5040	Benjamin Roberds	13.0	NaN	Drama Horror Thriller	Eva Boehnke	A Plague So Pleasant	38	
5041	Daniel Hsia	14.0	10443.0	Comedy Drama Romance	Alan Ruck	Shanghai Calling	1255	
5042	Jon Gunn	43.0	85222.0	Documentary	John August	My Date with Drew	4285	

5043 rows × 13 columns





# Drop the unnecessary rows using columns with high Null percentages(greater than 5%) of Null Values.

```
In [378... round(movies.isnull().sum().sort_values(ascending=False)/len(movies)*100,2)
```

```
Out[378]: gross                17.53  
budget                9.76  
title_year            2.14  
director_name         2.06  
num_critic_for_reviews 0.99  
num_user_for_reviews   0.40  
language              0.24  
actor_1_name          0.14  
genres                0.00  
movie_title           0.00  
num_voted_users        0.00  
imdb_score             0.00  
movie_facebook_likes   0.00  
dtype: float64
```

```
In [379... movies = movies[movies['gross'].notnull()]  
movies = movies[movies['budget'].notnull()]
```

```
In [380... round(movies.isnull().sum().sort_values(ascending=False)/len(movies)*100,2)
```

```
Out[380]: actor_1_name      0.08
          language        0.08
          num_critic_for_reviews  0.03
          director_name      0.00
          gross             0.00
          genres             0.00
          movie_title        0.00
          num_voted_users     0.00
          num_user_for_reviews  0.00
          budget            0.00
          title_year         0.00
          imdb_score         0.00
          movie_facebook_likes  0.00
          dtype: float64
```

**Drop more unnecessary rows which might have greater than five NaN values. Such rows are not of much use for the analysis and hence, should be removed.**

```
In [381... (movies.isnull().sum(axis=1).sort_values(ascending=False) > 5).sum()
```

```
Out[381]: 0
```

```
In [382... movies = movies[movies.isnull().sum(axis=1).sort_values(ascending=False) <= 5]
movies
```

C:\Users\HP\AppData\Local\Temp\ipykernel\_14236\820605845.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.

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

Out[382]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_
0	James Cameron	723.0	760505847.0	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	
1	Gore Verbinski	302.0	309404152.0	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End	471220	
2	Sam Mendes	602.0	200074175.0	Action Adventure Thriller	Christoph Waltz	Spectre	275868	
3	Christopher Nolan	813.0	448130642.0	Action Thriller	Tom Hardy	The Dark Knight Rises	1144337	
5	Andrew Stanton	462.0	73058679.0	Action Adventure Sci-Fi	Daryl Sabara	John Carter	212204	
...	...	...	...	...	...	...	...	...
5033	Shane Carruth	143.0	424760.0	Drama Sci-Fi Thriller	Shane Carruth	Primer	72639	
5034	Neill Dela Llane	35.0	70071.0	Thriller	Ian Gamazon	Cavite	589	
5035	Robert Rodriguez	56.0	2040920.0	Action Crime Drama Romance Thriller	Carlos Gallardo	El Mariachi	52055	
5037	Edward Burns	14.0	4584.0	Comedy Drama	Kerry Bishé	Newlyweds	1338	
5042	Jon Gunn	43.0	85222.0	Documentary	John August	My Date with Drew	4285	

3891 rows × 13 columns



Replace all the missing NaN values with 'English' in the Language column.

```
In [383... round(movies.isnull().sum().sort_values(ascending=False)/len(movies)*100,2)
```

```
Out[383]: actor_1_name      0.08  
language      0.08  
num_critic_for_reviews  0.03  
director_name  0.00  
gross         0.00  
genres        0.00  
movie_title   0.00  
num_voted_users  0.00  
num_user_for_reviews  0.00  
budget        0.00  
title_year    0.00  
imdb_score    0.00  
movie_facebook_likes  0.00  
dtype: float64
```

```
In [384... movies.language.describe()
```

```
Out[384]: count      3888  
unique         38  
top      English  
freq         3707  
Name: language, dtype: object
```

```
In [385... movies.language = movies.language.fillna('English')
```

```
In [386... round(movies.isnull().sum().sort_values(ascending=False)/len(movies)*100,2)
```

```
Out[386]: actor_1_name      0.08  
          num_critic_for_reviews  0.03  
          director_name      0.00  
          gross              0.00  
          genres             0.00  
          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
```

## Checking the percentage of the rows retained after completing all the tasks above.

```
In [387... movies
```

Out[387]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_
0	James Cameron	723.0	760505847.0	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	
1	Gore Verbinski	302.0	309404152.0	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End	471220	
2	Sam Mendes	602.0	200074175.0	Action Adventure Thriller	Christoph Waltz	Spectre	275868	
3	Christopher Nolan	813.0	448130642.0	Action Thriller	Tom Hardy	The Dark Knight Rises	1144337	
5	Andrew Stanton	462.0	73058679.0	Action Adventure Sci-Fi	Daryl Sabara	John Carter	212204	
...	...	...	...	...	...	...	...	...
5033	Shane Carruth	143.0	424760.0	Drama Sci-Fi Thriller	Shane Carruth	Primer	72639	
5034	Neill Dela Llane	35.0	70071.0	Thriller	Ian Gamazon	Cavite	589	
5035	Robert Rodriguez	56.0	2040920.0	Action Crime Drama Romance Thriller	Carlos Gallardo	El Mariachi	52055	
5037	Edward Burns	14.0	4584.0	Comedy Drama	Kerry Bishé	Newlyweds	1338	
5042	Jon Gunn	43.0	85222.0	Documentary	John August	My Date with Drew	4285	

3891 rows × 13 columns

In [388]: `len(movies)/len(originaldata)* 100`

Out[388]: 77.15645449137418

# Analysing Data - Converting the unit of budget and gross columns to million dollar.

```
In [389... movies['budget'] = movies['budget']/1000000  
movies['gross'] = movies['gross']/1000000
```

```
In [390... movies
```

Out[390]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_f
0	James Cameron	723.0	760.505847	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	
1	Gore Verbinski	302.0	309.404152	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End	471220	
2	Sam Mendes	602.0	200.074175	Action Adventure Thriller	Christoph Waltz	Spectre	275868	
3	Christopher Nolan	813.0	448.130642	Action Thriller	Tom Hardy	The Dark Knight Rises	1144337	
5	Andrew Stanton	462.0	73.058679	Action Adventure Sci-Fi	Daryl Sabara	John Carter	212204	
...	...	...	...	...	...	...	...	...
5033	Shane Carruth	143.0	0.424760	Drama Sci-Fi Thriller	Shane Carruth	Primer	72639	
5034	Neill Dela Llane	35.0	0.070071	Thriller	Ian Gamazon	Cavite	589	
5035	Robert Rodriguez	56.0	2.040920	Action Crime Drama Romance Thriller	Carlos Gallardo	El Mariachi	52055	
5037	Edward Burns	14.0	0.004584	Comedy Drama	Kerry Bishé	Newlyweds	1338	
5042	Jon Gunn	43.0	0.085222	Documentary	John August	My Date with Drew	4285	

3891 rows × 13 columns

Finding the movies with highest profit and store the top ten movies with highest profit in a new dataframe- top 10



```
In [391... movies['profit'] = movies['gross'] - movies['budget']
movies
```

Out[391]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_fo
0	James Cameron	723.0	760.505847	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	
1	Gore Verbinski	302.0	309.404152	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End	471220	
2	Sam Mendes	602.0	200.074175	Action Adventure Thriller	Christoph Waltz	Spectre	275868	
3	Christopher Nolan	813.0	448.130642	Action Thriller	Tom Hardy	The Dark Knight Rises	1144337	
5	Andrew Stanton	462.0	73.058679	Action Adventure Sci-Fi	Daryl Sabara	John Carter	212204	
...	...	...	...	...	...	...	...	
5033	Shane Carruth	143.0	0.424760	Drama Sci-Fi Thriller	Shane Carruth	Primer	72639	
5034	Neill Dela Llana	35.0	0.070071	Thriller	Ian Gamazon	Cavite	589	
5035	Robert Rodriguez	56.0	2.040920	Action Crime Drama Romance Thriller	Carlos Gallardo	El Mariachi	52055	
5037	Edward Burns	14.0	0.004584	Comedy Drama	Kerry Bishé	Newlyweds	1338	
5042	Jon Gunn	43.0	0.085222	Documentary	John August	My Date with Drew	4285	

3891 rows × 14 columns

```
In [392... movies.sort_values(by='profit',ascending=False)
```

Out[392]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user
0	James Cameron	723.0	760.505847	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	
29	Colin Trevorrow	644.0	652.177271	Action Adventure Sci-Fi Thriller	Bryce Dallas Howard	Jurassic World	418214	
26	James Cameron	315.0	658.672302	Drama Romance	Leonardo DiCaprio	Titanic	793059	
3024	George Lucas	282.0	460.935665	Action Adventure Fantasy Sci-Fi	Harrison Ford	Star Wars: Episode IV - A New Hope	911097	
3080	Steven Spielberg	215.0	434.949459	Family Sci-Fi	Henry Thomas	E.T. the Extra-Terrestrial	281842	
...	...	...	...	...	...	...	...	
2334	Katsuhiro Ôtomo	105.0	0.410388	Action Adventure Animation Family Sci-Fi Thriller	William Hootkins	Steamboy	13727	
2323	Hayao Miyazaki	174.0	2.298191	Adventure Animation Fantasy	Minnie Driver	Princess Mononoke	221552	
3005	Lajos Koltai	73.0	0.195888	Drama Romance War	Marcell Nagy	Fateless	5603	
3859	Chan-wook Park	202.0	0.211667	Crime Drama	Min-sik Choi	Lady Vengeance	53508	
2988	Joon-ho Bong	363.0	2.201412	Comedy Drama Horror Sci-Fi	Doona Bae	The Host	68883	

3891 rows × 14 columns



```
In [393... top10 = movies.sort_values(by='profit',ascending=False).head(10)
top10
```

Out[393]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_u
0	James Cameron	723.0	760.505847	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	
29	Colin Trevorrow	644.0	652.177271	Action Adventure Sci-Fi Thriller	Bryce Dallas Howard	Jurassic World	418214	
26	James Cameron	315.0	658.672302	Drama Romance	Leonardo DiCaprio	Titanic	793059	
3024	George Lucas	282.0	460.935665	Action Adventure Fantasy Sci-Fi	Harrison Ford	Star Wars: Episode IV - A New Hope	911097	
3080	Steven Spielberg	215.0	434.949459	Family Sci-Fi	Henry Thomas	E.T. the Extra-Terrestrial	281842	
794	Joss Whedon	703.0	623.279547	Action Adventure Sci-Fi	Chris Hemsworth	The Avengers	995415	
17	Joss Whedon	703.0	623.279547	Action Adventure Sci-Fi	Chris Hemsworth	The Avengers	995415	
509	Roger Allers	186.0	422.783777	Adventure Animation Drama Family Musical	Matthew Broderick	The Lion King	644348	
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	



## Drop the duplicate values.

In [394... `movies.drop_duplicates(keep='first', inplace=True)`

In [395... `movies`

Out[395]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_fo
0	James Cameron	723.0	760.505847	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	
1	Gore Verbinski	302.0	309.404152	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End	471220	
2	Sam Mendes	602.0	200.074175	Action Adventure Thriller	Christoph Waltz	Spectre	275868	
3	Christopher Nolan	813.0	448.130642	Action Thriller	Tom Hardy	The Dark Knight Rises	1144337	
5	Andrew Stanton	462.0	73.058679	Action Adventure Sci-Fi	Daryl Sabara	John Carter	212204	
...	...	...	...	...	...	...	...	
5033	Shane Carruth	143.0	0.424760	Drama Sci-Fi Thriller	Shane Carruth	Primer	72639	
5034	Neill Dela Llana	35.0	0.070071	Thriller	Ian Gamazon	Cavite	589	
5035	Robert Rodriguez	56.0	2.040920	Action Crime Drama Romance Thriller	Carlos Gallardo	El Mariachi	52055	
5037	Edward Burns	14.0	0.004584	Comedy Drama	Kerry Bishé	Newlyweds	1338	
5042	Jon Gunn	43.0	0.085222	Documentary	John August	My Date with Drew	4285	

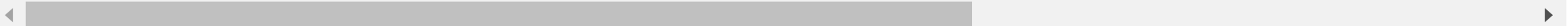
3856 rows × 14 columns



```
In [396... top10 = movies.sort_values(by='profit',ascending=False).head(10)
top10
```

Out[396]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_u
0	James Cameron	723.0	760.505847	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	
29	Colin Trevorrow	644.0	652.177271	Action Adventure Sci-Fi Thriller	Bryce Dallas Howard	Jurassic World	418214	
26	James Cameron	315.0	658.672302	Drama Romance	Leonardo DiCaprio	Titanic	793059	
3024	George Lucas	282.0	460.935665	Action Adventure Fantasy Sci-Fi	Harrison Ford	Star Wars: Episode IV - A New Hope	911097	
3080	Steven Spielberg	215.0	434.949459	Family Sci-Fi	Henry Thomas	E.T. the Extra-Terrestrial	281842	
17	Joss Whedon	703.0	623.279547	Action Adventure Sci-Fi	Chris Hemsworth	The Avengers	995415	
509	Roger Allers	186.0	422.783777	Adventure Animation Drama Family Musical	Matthew Broderick	The Lion King	644348	
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	



**Finding IMDB top 100 movies with highest IMDB rating and number of voted users greater than 25000. Add a rank column containing the values 1 to 100.**

```
In [397... IMDb_Top_100 = movies[movies['num_voted_users'] > 25000].sort_values(by='imdb_score', ascending=False).head(100)
IMDb_Top_100
```

Out[397]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_fo
1937	Frank Darabont	199.0	28.341469	Crime Drama	Morgan Freeman	The Shawshank Redemption	1689764	
3466	Francis Ford Coppola	208.0	134.821952	Crime Drama	Al Pacino	The Godfather	1155770	
2837	Francis Ford Coppola	149.0	57.300000	Crime Drama	Robert De Niro	The Godfather: Part II	790926	
66	Christopher Nolan	645.0	533.316061	Action Crime Drama Thriller	Christian Bale	The Dark Knight	1676169	
4498	Sergio Leone	181.0	6.100000	Western	Clint Eastwood	The Good, the Bad and the Ugly	503509	
...	...	...	...	...	...	...	...	
27	Anthony Russo	516.0	407.197282	Action Adventure Sci-Fi	Robert Downey Jr.	Captain America: Civil War	272670	
4000	Juan José Campanella	262.0	20.167424	Drama Mystery Thriller	Ricardo Darín	The Secret in Their Eyes	131831	
3550	Denis Villeneuve	226.0	6.857096	Drama Mystery War	Lubna Azabal	Incendies	80429	
3970	Victor Fleming	157.0	198.655278	Drama History Romance War	Hattie McDaniel	Gone with the Wind	215340	
2047	Hayao Miyazaki	212.0	4.710455	Adventure Animation Family Fantasy	Christian Bale	Howl's Moving Castle	214091	

100 rows × 14 columns



In [398... IMDb\_Top\_100['Rank'] = IMDb\_Top\_100['imdb\_score'].rank(method='first', ascending=False)

IMDb\_Top\_100

Out[398]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_fo
<b>1937</b>	Frank Darabont	199.0	28.341469	Crime Drama	Morgan Freeman	The Shawshank Redemption	1689764	
<b>3466</b>	Francis Ford Coppola	208.0	134.821952	Crime Drama	Al Pacino	The Godfather	1155770	
<b>2837</b>	Francis Ford Coppola	149.0	57.300000	Crime Drama	Robert De Niro	The Godfather: Part II	790926	
<b>66</b>	Christopher Nolan	645.0	533.316061	Action Crime Drama Thriller	Christian Bale	The Dark Knight	1676169	
<b>4498</b>	Sergio Leone	181.0	6.100000	Western	Clint Eastwood	The Good, the Bad and the Ugly	503509	
...	...	...	...	...	...	...	...	
<b>27</b>	Anthony Russo	516.0	407.197282	Action Adventure Sci-Fi	Robert Downey Jr.	Captain America: Civil War	272670	
<b>4000</b>	Juan José Campanella	262.0	20.167424	Drama Mystery Thriller	Ricardo Darín	The Secret in Their Eyes	131831	
<b>3550</b>	Denis Villeneuve	226.0	6.857096	Drama Mystery War	Lubna Azabal	Incendies	80429	
<b>3970</b>	Victor Fleming	157.0	198.655278	Drama History Romance War	Hattie McDaniel	Gone with the Wind	215340	
<b>2047</b>	Hayao Miyazaki	212.0	4.710455	Adventure Animation Family Fantasy	Christian Bale	Howl's Moving Castle	214091	

100 rows × 15 columns





```
In [399... IMDb_Top_100[IMDb_Top_100['language']!='English']
```

Out[399]:

	director_name	num_critic_for_reviews	gross		genres	actor_1_name	movie_title	num_voted_users	num_user_foi
4498	Sergio Leone	181.0	6.100000		Western	Clint Eastwood	The Good, the Bad and the Ugly	503509	
4747	Akira Kurosawa	153.0	0.269061		Action Adventure Drama	Takashi Shimura	Seven Samurai	229012	
4029	Fernando Meirelles	214.0	7.563397		Crime Drama	Alice Braga	City of God	533200	
2373	Hayao Miyazaki	246.0	10.049886		Adventure Animation Family Fantasy	Bunta Sugawara	Spirited Away	417971	
4259	Florian Henckel von Donnersmarck	215.0	11.284657		Drama Thriller	Sebastian Koch	The Lives of Others	259379	
4921	Majid Majidi	46.0	0.925402		Drama Family	Bahare Seddiqi	Children of Heaven	27882	
2323	Hayao Miyazaki	174.0	2.298191		Adventure Animation Fantasy	Minnie Driver	Princess Mononoke	221552	
2970	Wolfgang Petersen	96.0	11.433134		Adventure Drama Thriller War	Jürgen Prochnow	Das Boot	168203	
4105	Chan-wook Park	305.0	2.181290		Drama Mystery Thriller	Min-sik Choi	Oldboy	356181	
4659	Asghar Farhadi	354.0	7.098492		Drama Mystery	Shahab Hosseini	A Separation	151812	
1329	S.S. Rajamouli	44.0	6.498000		Action Adventure Drama Fantasy War	Tamannaah Bhatia	Baahubali: The Beginning	62756	
1298	Jean-Pierre Jeunet	242.0	33.201661		Comedy Romance	Mathieu Kassovitz	Amélie	534262	
2734	Fritz Lang	260.0	0.026435		Drama Sci-Fi	Brigitte Helm	Metropolis	111841	
4033	Thomas Vinterberg	349.0	0.610968		Drama	Thomas Bo Larsen	The Hunt	170155	

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_fo
2829	Oliver Hirschbiegel	192.0	5.501940	Biography Drama History War	Thomas Kretschmann	Downfall	248354	
2551	Guillermo del Toro	406.0	37.623143	Drama Fantasy War	Ivana Baquero	Pan's Labyrinth	467234	
4000	Juan José Campanella	262.0	20.167424	Drama Mystery Thriller	Ricardo Darín	The Secret in Their Eyes	131831	
3550	Denis Villeneuve	226.0	6.857096	Drama Mystery War	Lubna Azabal	Incendies	80429	
2047	Hayao Miyazaki	212.0	4.710455	Adventure Animation Family Fantasy	Christian Bale	Howl's Moving Castle	214091	

## Finding the best directors with high imdb scores and store top 10 directors in a new coloumn.

```
In [400]: top10director = movies.groupby('director_name').imdb_score.mean().sort_values(ascending=False).head(10)
top10director
```

```
Out[400]: director_name
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
Name: imdb_score, dtype: float64
```

# Finding popular genres.

```
In [401]: TempGenre = movies.genres.str.split('|', expand=True).iloc[:,0:2]
TempGenre.columns=['genre_1', 'genre_2']
TempGenre.genre_2.fillna(TempGenre.genre_1, inplace=True)
TempGenre
```

```
Out[401]:
```

	genre_1	genre_2
0	Action	Adventure
1	Action	Adventure
2	Action	Adventure
3	Action	Thriller
5	Action	Adventure
...	...	...
5033	Drama	Sci-Fi
5034	Thriller	Thriller
5035	Action	Crime
5037	Comedy	Drama
5042	Documentary	Documentary

3856 rows × 2 columns

```
In [402]: movies = pd.concat([movies, TempGenre], axis=1)
movies
```

Out[402]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_f
0	James Cameron	723.0	760.505847	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	
1	Gore Verbinski	302.0	309.404152	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End	471220	
2	Sam Mendes	602.0	200.074175	Action Adventure Thriller	Christoph Waltz	Spectre	275868	
3	Christopher Nolan	813.0	448.130642	Action Thriller	Tom Hardy	The Dark Knight Rises	1144337	
5	Andrew Stanton	462.0	73.058679	Action Adventure Sci-Fi	Daryl Sabara	John Carter	212204	
...	...	...	...	...	...	...	...	...
5033	Shane Carruth	143.0	0.424760	Drama Sci-Fi Thriller	Shane Carruth	Primer	72639	
5034	Neill Dela Llane	35.0	0.070071	Thriller	Ian Gamazon	Cavite	589	
5035	Robert Rodriguez	56.0	2.040920	Action Crime Drama Romance Thriller	Carlos Gallardo	El Mariachi	52055	
5037	Edward Burns	14.0	0.004584	Comedy Drama	Kerry Bishé	Newlyweds	1338	
5042	Jon Gunn	43.0	0.085222	Documentary	John August	My Date with Drew	4285	

3856 rows × 16 columns


In [403... `movies.groupby(['genre_1', 'genre_2']).gross.mean().sort_values(ascending=False).head(5)`

```
Out[403]: genre_1    genre_2
Family      Sci-Fi      434.949459
Adventure   Sci-Fi      228.627758
            Family      118.919540
            Animation    116.998550
Action      Adventure    109.595465
Name: gross, dtype: float64
```

## Finding the critic-favourite and audience-favourite actors.

```
In [404... Meryl_Streep = movies[movies['actor_1_name']=='Meryl Streep']
Leo_Caprio = movies[movies['actor_1_name']=='Leonardo DiCaprio']
Brad_Pitt = movies[movies['actor_1_name']=='Brad Pitt']
```

```
In [405... Combined = Meryl_Streep.append([Leo_Caprio,Brad_Pitt])
Combined
```

```
C:\Users\HP\AppData\Local\Temp\ipykernel_14236\3117084833.py:1: FutureWarning: The frame.append method is deprecated and will
be removed from pandas in a future version. Use pandas.concat instead.
    Combined = Meryl_Streep.append([Leo_Caprio,Brad_Pitt])
```

Out[405]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_use
410	Nancy Meyers	187.0	112.703470	Comedy Drama Romance	Meryl Streep	It's Complicated	6986
1106	Curtis Hanson	42.0	46.815748	Action Adventure Crime Thriller	Meryl Streep	The River Wild	3254
1204	Nora Ephron	252.0	94.125426	Biography Drama Romance	Meryl Streep	Julie & Julia	7926
1408	David Frankel	208.0	124.732962	Comedy Drama Romance	Meryl Streep	The Devil Wears Prada	28617
1483	Robert Redford	227.0	14.998070	Drama Thriller War	Meryl Streep	Lions for Lambs	4117
1575	Sydney Pollack	66.0	87.100000	Biography Drama Romance	Meryl Streep	Out of Africa	5235
1618	David Frankel	234.0	63.536011	Comedy Drama Romance	Meryl Streep	Hope Springs	3429
1674	Carl Franklin	64.0	23.209440	Drama	Meryl Streep	One True Thing	928
1925	Stephen Daldry	174.0	41.597830	Drama Romance	Meryl Streep	The Hours	10212
2781	Phyllida Lloyd	331.0	29.959436	Biography Drama History	Meryl Streep	The Iron Lady	8232
3135	Robert Altman	211.0	20.338609	Comedy Drama Music	Meryl Streep	A Prairie Home Companion	1965
26	James Cameron	315.0	658.672302	Drama Romance	Leonardo DiCaprio	Titanic	79305
50	Baz Luhrmann	490.0	144.812796	Drama Romance	Leonardo DiCaprio	The Great Gatsby	36297
97	Christopher Nolan	642.0	292.568851	Action Adventure Sci-Fi Thriller	Leonardo DiCaprio	Inception	146820
179	Alejandro G. Iñárritu	556.0	183.635922	Adventure Drama Thriller Western	Leonardo DiCaprio	The Revenant	40602
257	Martin Scorsese	267.0	102.608827	Biography Drama	Leonardo DiCaprio	The Aviator	26437

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_use
296	Quentin Tarantino	765.0	162.804648	Drama Western	Leonardo DiCaprio	Django Unchained	95511
307	Edward Zwick	166.0	57.366262	Adventure Drama Thriller	Leonardo DiCaprio	Blood Diamond	40029
308	Martin Scorsese	606.0	116.866727	Biography Comedy Crime Drama	Leonardo DiCaprio	The Wolf of Wall Street	78058
326	Martin Scorsese	233.0	77.679638	Crime Drama	Leonardo DiCaprio	Gangs of New York	31403
361	Martin Scorsese	352.0	132.373442	Crime Drama Thriller	Leonardo DiCaprio	The Departed	87364
452	Martin Scorsese	490.0	127.968405	Mystery Thriller	Leonardo DiCaprio	Shutter Island	78609
641	Ridley Scott	238.0	39.380442	Action Drama Thriller	Leonardo DiCaprio	Body of Lies	17424
911	Steven Spielberg	194.0	164.435221	Biography Crime Drama	Leonardo DiCaprio	Catch Me If You Can	52580
990	Danny Boyle	118.0	39.778599	Adventure Drama Thriller	Leonardo DiCaprio	The Beach	17616
1114	Sam Mendes	323.0	22.877808	Drama Romance	Leonardo DiCaprio	Revolutionary Road	15259
1422	Randall Wallace	83.0	56.876365	Action Adventure	Leonardo DiCaprio	The Man in the Iron Mask	12527
1453	Clint Eastwood	392.0	37.304950	Biography Crime Drama	Leonardo DiCaprio	J. Edgar	10272
1560	Sam Raimi	63.0	18.636537	Action Thriller Western	Leonardo DiCaprio	The Quick and the Dead	6919
2067	Jerry Zaks	45.0	12.782508	Drama	Leonardo DiCaprio	Marvin's Room	2016
2757	Baz Luhrmann	106.0	46.338728	Drama Romance	Leonardo DiCaprio	Romeo + Juliet	16779



	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_use
3476	Baz Luhrmann	490.0	144.812796	Drama Romance	Leonardo DiCaprio	The Great Gatsby	36295
101	David Fincher	362.0	127.490802	Drama Fantasy Romance	Brad Pitt	The Curious Case of Benjamin Button	45934
147	Wolfgang Petersen	220.0	133.228348	Adventure	Brad Pitt	Troy	38167
254	Steven Soderbergh	198.0	125.531634	Crime Thriller	Brad Pitt	Ocean's Twelve	28485
255	Doug Liman	233.0	186.336103	Action Comedy Crime Romance Thriller	Brad Pitt	Mr. & Mrs. Smith	34886
382	Tony Scott	142.0	0.026871	Action Crime Thriller	Brad Pitt	Spy Game	12125
400	Steven Soderbergh	186.0	183.405771	Crime Thriller	Brad Pitt	Ocean's Eleven	40264
470	David Ayer	406.0	85.707116	Action Drama War	Brad Pitt	Fury	30318
611	Jean-Jacques Annaud	76.0	37.901509	Adventure Biography Drama History War	Brad Pitt	Seven Years in Tibet	9638
683	David Fincher	315.0	37.023395	Drama	Brad Pitt	Fight Club	134746
792	Patrick Gilmore	98.0	26.288320	Adventure Animation Comedy Drama Family Fantas...	Brad Pitt	Sinbad: Legend of the Seven Seas	3614
940	Neil Jordan	120.0	105.264608	Drama Fantasy Horror	Brad Pitt	Interview with the Vampire: The Vampire Chroni...	23975
1490	Terrence Malick	584.0	13.303319	Drama Fantasy	Brad Pitt	The Tree of Life	13636

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_use
1722	Andrew Dominik	273.0	3.904982	Biography Crime Drama History Western	Brad Pitt	The Assassination of Jesse James by the Coward...	13610
2204	Alejandro G. Iñárritu	285.0	34.300771	Drama	Brad Pitt	Babel	24379
2333	Angelina Jolie Pitt	131.0	0.531009	Drama Romance	Brad Pitt	By the Sea	797
2682	Andrew Dominik	414.0	14.938570	Crime Thriller	Brad Pitt	Killing Them Softly	11162
2898	Tony Scott	122.0	12.281500	Action Crime Drama Romance Thriller	Brad Pitt	True Romance	16349



In [406... Combined.groupby('actor\_1\_name').num\_critic\_for\_reviews.mean()

```
Out[406]: actor_1_name
Brad Pitt      245.000000
Leonardo DiCaprio  330.190476
Meryl Streep    181.454545
Name: num_critic_for_reviews, dtype: float64
```

In [407... Combined

Out[407]:

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_use
410	Nancy Meyers	187.0	112.703470	Comedy Drama Romance	Meryl Streep	It's Complicated	6986
1106	Curtis Hanson	42.0	46.815748	Action Adventure Crime Thriller	Meryl Streep	The River Wild	3254
1204	Nora Ephron	252.0	94.125426	Biography Drama Romance	Meryl Streep	Julie & Julia	7926
1408	David Frankel	208.0	124.732962	Comedy Drama Romance	Meryl Streep	The Devil Wears Prada	28617
1483	Robert Redford	227.0	14.998070	Drama Thriller War	Meryl Streep	Lions for Lambs	4117
1575	Sydney Pollack	66.0	87.100000	Biography Drama Romance	Meryl Streep	Out of Africa	5235
1618	David Frankel	234.0	63.536011	Comedy Drama Romance	Meryl Streep	Hope Springs	3429
1674	Carl Franklin	64.0	23.209440	Drama	Meryl Streep	One True Thing	928
1925	Stephen Daldry	174.0	41.597830	Drama Romance	Meryl Streep	The Hours	10212
2781	Phyllida Lloyd	331.0	29.959436	Biography Drama History	Meryl Streep	The Iron Lady	8232
3135	Robert Altman	211.0	20.338609	Comedy Drama Music	Meryl Streep	A Prairie Home Companion	1965
26	James Cameron	315.0	658.672302	Drama Romance	Leonardo DiCaprio	Titanic	79305
50	Baz Luhrmann	490.0	144.812796	Drama Romance	Leonardo DiCaprio	The Great Gatsby	36297
97	Christopher Nolan	642.0	292.568851	Action Adventure Sci-Fi Thriller	Leonardo DiCaprio	Inception	146820
179	Alejandro G. Iñárritu	556.0	183.635922	Adventure Drama Thriller Western	Leonardo DiCaprio	The Revenant	40602
257	Martin Scorsese	267.0	102.608827	Biography Drama	Leonardo DiCaprio	The Aviator	26437

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_use
296	Quentin Tarantino	765.0	162.804648	Drama Western	Leonardo DiCaprio	Django Unchained	95511
307	Edward Zwick	166.0	57.366262	Adventure Drama Thriller	Leonardo DiCaprio	Blood Diamond	40029
308	Martin Scorsese	606.0	116.866727	Biography Comedy Crime Drama	Leonardo DiCaprio	The Wolf of Wall Street	78058
326	Martin Scorsese	233.0	77.679638	Crime Drama	Leonardo DiCaprio	Gangs of New York	31403
361	Martin Scorsese	352.0	132.373442	Crime Drama Thriller	Leonardo DiCaprio	The Departed	87364
452	Martin Scorsese	490.0	127.968405	Mystery Thriller	Leonardo DiCaprio	Shutter Island	78609
641	Ridley Scott	238.0	39.380442	Action Drama Thriller	Leonardo DiCaprio	Body of Lies	17424
911	Steven Spielberg	194.0	164.435221	Biography Crime Drama	Leonardo DiCaprio	Catch Me If You Can	52580
990	Danny Boyle	118.0	39.778599	Adventure Drama Thriller	Leonardo DiCaprio	The Beach	17616
1114	Sam Mendes	323.0	22.877808	Drama Romance	Leonardo DiCaprio	Revolutionary Road	15259
1422	Randall Wallace	83.0	56.876365	Action Adventure	Leonardo DiCaprio	The Man in the Iron Mask	12527
1453	Clint Eastwood	392.0	37.304950	Biography Crime Drama	Leonardo DiCaprio	J. Edgar	10272
1560	Sam Raimi	63.0	18.636537	Action Thriller Western	Leonardo DiCaprio	The Quick and the Dead	6919
2067	Jerry Zaks	45.0	12.782508	Drama	Leonardo DiCaprio	Marvin's Room	2016
2757	Baz Luhrmann	106.0	46.338728	Drama Romance	Leonardo DiCaprio	Romeo + Juliet	16779

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_use
3476	Baz Luhrmann	490.0	144.812796	Drama Romance	Leonardo DiCaprio	The Great Gatsby	36295
101	David Fincher	362.0	127.490802	Drama Fantasy Romance	Brad Pitt	The Curious Case of Benjamin Button	45934
147	Wolfgang Petersen	220.0	133.228348	Adventure	Brad Pitt	Troy	38167
254	Steven Soderbergh	198.0	125.531634	Crime Thriller	Brad Pitt	Ocean's Twelve	28485
255	Doug Liman	233.0	186.336103	Action Comedy Crime Romance Thriller	Brad Pitt	Mr. & Mrs. Smith	34886
382	Tony Scott	142.0	0.026871	Action Crime Thriller	Brad Pitt	Spy Game	12125
400	Steven Soderbergh	186.0	183.405771	Crime Thriller	Brad Pitt	Ocean's Eleven	40264
470	David Ayer	406.0	85.707116	Action Drama War	Brad Pitt	Fury	30318
611	Jean-Jacques Annaud	76.0	37.901509	Adventure Biography Drama History War	Brad Pitt	Seven Years in Tibet	9638
683	David Fincher	315.0	37.023395	Drama	Brad Pitt	Fight Club	134746
792	Patrick Gilmore	98.0	26.288320	Adventure Animation Comedy Drama Family Fantas...	Brad Pitt	Sinbad: Legend of the Seven Seas	3614
940	Neil Jordan	120.0	105.264608	Drama Fantasy Horror	Brad Pitt	Interview with the Vampire: The Vampire Chroni...	23975
1490	Terrence Malick	584.0	13.303319	Drama Fantasy	Brad Pitt	The Tree of Life	13636

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	num_voted_use
1722	Andrew Dominik	273.0	3.904982	Biography Crime Drama History Western	Brad Pitt	The Assassination of Jesse James by the Coward...	13610
2204	Alejandro G. Iñárritu	285.0	34.300771	Drama	Brad Pitt	Babel	24379
2333	Angelina Jolie Pitt	131.0	0.531009	Drama Romance	Brad Pitt	By the Sea	791
2682	Andrew Dominik	414.0	14.938570	Crime Thriller	Brad Pitt	Killing Them Softly	11162
2898	Tony Scott	122.0	12.281500	Action Crime Drama Romance Thriller	Brad Pitt	True Romance	16349



```
In [408... Combined.num_user_for_reviews = Combined.num_user_for_reviews.astype('int')
Combined.num_user_for_reviews
```

```
Out[408]: 410      214
          1106      69
          1204      277
          1408      631
          1483      298
          1575      200
          1618      178
          1674      112
          1925      660
          2781      350
          3135      280
           26     2528
           50      753
           97     2803
          179     1188
          257      799
          296     1193
          307      657
          308     1138
          326     1166
          361     2054
          452      964
          641      263
          911      667
          990      548
         1114      414
         1422      244
         1453      279
         1560      216
         2067       71
         2757      506
         3476      753
          101      822
          147     1694
          254      627
          255      798
          382      361
          400      845
          470      701
          611      119
```

```

683      2968
792        91
940       406
1490      975
1722      415
2204      908
2333        61
2682       369
2898       460

```

Name: num\_user\_for\_reviews, dtype: int32

```
In [409... Combined.groupby('actor_1_name').num_user_for_reviews.mean()
```

```

Out[409]: actor_1_name
Brad Pitt      742.352941
Leonardo DiCaprio  914.476190
Meryl Streep    297.181818
Name: num_user_for_reviews, dtype: float64

```

```
In [410... Combined.groupby('actor_1_name')[['num_critic_for_reviews','num_user_for_reviews']].mean()
```

```

Out[410]:

```

	num_critic_for_reviews	num_user_for_reviews
<b>actor_1_name</b>		
<b>Brad Pitt</b>	245.000000	742.352941
<b>Leonardo DiCaprio</b>	330.190476	914.476190
<b>Meryl Streep</b>	181.454545	297.181818