In [141…

```
# Supress Warnings

import warnings
warnings.filterwarnings('ignore')
```

In [243…

```
# Import the numpy and pandas packages

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

This projects is based on the IMDB dataset and we have find some insights out of the Data. Steps going to be taken:

1. View the data and Understand
2. Cleaning the DATA
3. Make Changes in the Dataset for further analysis.
4. Performing EDA on the DataSet
5. Visualising the Results using a Graph or a Chart
6. Providng Conclusions

Task 1: Reading and Inspection

- ### Subtask 1.1: Import and read

Import and read the movie database. Store it in a variable called `movies` .

In [143…

```
movies = pd.read_csv(r"C:\Users\eva\Desktop\assignements\imdb\IMDB_Movies.csv")
movies
```

Out[143…

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_
0	Color	James Cameron	723.0	178.0	0.0	855.0	Joel
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0	O
2	Color	Sam Mendes	602.0	148.0	0.0	161.0	Rory K
3	Color	Christopher Nolan	813.0	164.0	22000.0	23000.0	Christia
4	NaN	Doug Walker	NaN	NaN	131.0	NaN	Rob \
...	...	...	...	...	...	...	
5038	Color	Scott Smith	1.0	87.0	2.0	318.0	D
5039	Color	NaN	43.0	43.0	NaN	319.0	Valorie
5040	Color	Benjamin Roberds	13.0	76.0	0.0	0.0	M
5041	Color	Daniel Hsia	14.0	100.0	0.0	489.0	Daniel H

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_
5042	Color	Jon Gunn	43.0	90.0	16.0	16.0	Her:

5043 rows × 28 columns

- ### Subtask 1.2: Inspect the dataframe

Inspect the dataframe's columns, shapes, variable types etc.

In [144...

```
print('Columns')
print(movies.columns)

print('Shapes')

print(movies.shape)

print('Variable_types')

print(movies.dtypes)
```

```
Columns
Index(['color', 'director_name', 'num_critic_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'],
      dtype='object')
Shapes
(5043, 28)
Variable_types
color                         object
director_name                 object
num_critic_for_reviews       float64
duration                     float64
director_facebook_likes      float64
actor_3_facebook_likes       float64
actor_2_name                  object
actor_1_facebook_likes       float64
gross                        float64
genres                        object
actor_1_name                  object
movie_title                   object
num_voted_users                int64
cast_total_facebook_likes      int64
actor_3_name                  object
facenumber_in_poster         float64
plot_keywords                 object
movie_imdb_link               object
num_user_for_reviews          object
language                      object
country                       object
content_rating                object
budget                       float64
title_year                   float64
actor_2_facebook_likes       float64
imdb_score                   float64
aspect_ratio                 float64
```

```
movie_facebook_likes          int64
dtype: object
```

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.

In [145…

```
print(movies.isnull().sum(axis = 0))

# Table contains lot of Null Values most of them accumulated in certain columns
```

```
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          20
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
```

In [146…

```
print(movies.isnull().sum(axis = 1))
```

```
0        0
1        0
2        0
3        0
4       13
        ..
5038     4
5039     5
5040     4
5041     2
5042     0
Length: 5043, dtype: int64
```

In [147…

```
Null_by_coloumns = movies.isnull().sum(axis = 0)
Null_by_coloumns = Null_by_coloumns.to_frame()
Null_by_coloumns = Null_by_coloumns.rename(columns= {0: 'Num_of_Nulls'})
```

```
Total_null_values = Null_by_coloumns.Num_of_Nulls.sum()
print(Total_null_values)
Null_by_coloumns['Null_per'] = round(((Null_by_coloumns.Num_of_Nulls/Total_null_values)*1(
print(Null_by_coloumns)

# Calculated the percentage of Null Values in each coloumn for better understanding
```

```
2697
                           Num_of_Nulls  Null_per
color                                19      0.70
director_name                       104      3.86
num_critic_for_reviews               50      1.85
duration                             15      0.56
director_facebook_likes             104      3.86
actor_3_facebook_likes               23      0.85
actor_2_name                         13      0.48
actor_1_facebook_likes                7      0.26
gross                               884     32.78
genres                                0      0.00
actor_1_name                          7      0.26
movie_title                           0      0.00
num_voted_users                       0      0.00
cast_total_facebook_likes             0      0.00
actor_3_name                         23      0.85
facenumber_in_poster                 13      0.48
plot_keywords                       153      5.67
movie_imdb_link                       0      0.00
num_user_for_reviews                 20      0.74
language                             12      0.44
country                               5      0.19
content_rating                      303     11.23
budget                              492     18.24
title_year                          108      4.00
actor_2_facebook_likes               13      0.48
imdb_score                            0      0.00
aspect_ratio                        329     12.20
movie_facebook_likes                  0      0.00
```

- ### 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

In [148…

```
movies_new = movies.drop(['color','director_facebook_likes', 'actor_1_facebook_likes', 'ac
movies_new
```

Out[148…

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

5043 rows × 13 columns

- ### Subtask 2.3: Drop unecessary rows using columns with high Null percentages

Now, 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.

In [149…

```
Greater_5per_columns = set(Null_by_coloumns.loc[(Null_by_coloumns['Null_per'] > 5)].index)
movies_new_columns = set(movies_new.columns)
Greater_5per_columns
movies_new_columns

intersection = list(Greater_5per_columns & movies_new_columns)
intersection

actual_data = movies_new.dropna(subset = intersection)
actual_data
```

```
actual_data.language.isnull().sum()

# Dropping Null Values for Coloumns that have high Null Value percentage
```

Out[149…

```
3
```

- ### Subtask 2.4: 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'`.

In [150…

```
pd.options.mode.chained_assignment = None
# (To disable the copy warning)

actual_data['language'].fillna("English", inplace = True)
print(actual_data)
actual_data.language.isnull().sum()

# Hadling some of the Null Values by either replacing or deleting
```

```
          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  English  237000000.0      2009.0         7.9
1                    1238  English  300000000.0      2007.0         7.1
```

```
2                       994  English  245000000.0      2015.0          6.8
3                      2701  English  250000000.0      2012.0          8.5
5                       738  English  263700000.0      2012.0          6.6
...                     ...      ...          ...         ...          ...
5033                    371  English       7000.0      2004.0          7.0
5034                     35  English       7000.0      2005.0          6.3
5035                    130  Spanish       7000.0      1992.0          6.9
5037                     14  English       9000.0      2011.0          6.4
5042                     84  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]
```

Out[150…

```
0
```

- ### Subtask 2.5: 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.

In [151…

```
print(len(actual_data.index))
print((len(actual_data.index)/len(movies.index))*100)
```

```
3891
77.15645449137418
```

Checkpoint 1: You might have noticed that we still have around `77%` of the rows!

Task 3: Data Analysis

In this Part We will be performing most of the detailed Analysis

- ### Subtask 3.1: Change the unit of columns

Convert the unit of the `budget` and `gross` columns from `$` to `million $`.

In [152…

```
actual_data.budget = round(actual_data["budget"]/1000000,2)
actual_data.gross = round(actual_data["gross"]/1000000,2)
print(actual_data)

# Changing the Unit of Gross and Budget Fields so that we can read the data easily
```

```
         director_name  num_critic_for_reviews   gross  \
0        James Cameron                   723.0  760.51
1       Gore Verbinski                   302.0  309.40
2           Sam Mendes                   602.0  200.07
3    Christopher Nolan                   813.0  448.13
5       Andrew Stanton                   462.0   73.06
```

```
...                   ...                          ...     ...
5033      Shane Carruth                          143.0    0.42
5034   Neill Dela Llana                           35.0    0.07
5035   Robert Rodriguez                           56.0    2.04
5037       Edward Burns                           14.0    0.00
5042           Jon Gunn                           43.0    0.09

                                   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  English  237.00      2009.0         7.9
1                    1238  English  300.00      2007.0         7.1
2                     994  English  245.00      2015.0         6.8
3                    2701  English  250.00      2012.0         8.5
5                     738  English  263.70      2012.0         6.6
...                   ...      ...     ...         ...         ...
5033                  371  English    0.01      2004.0         7.0
5034                   35  English    0.01      2005.0         6.3
5035                  130  Spanish    0.01      1992.0         6.9
5037                   14  English    0.01      2011.0         6.4
5042                   84  English    0.00      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. Plot `profit` (y-axis) vs `budget` (x- axis) and observe the outliers using the appropriate chart type.
4. Extract the top ten profiting movies in descending order and store them in a new dataframe - `top10`

In [153…

```
actual_data['profit']= actual_data.gross-actual_data.budget
actual_data
```

Out[153…

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	nu
0	James Cameron	723.0	760.51	Action|Adventure|Fantasy|Sci-Fi	CCH Pounder	Avatar	
1	Gore Verbinski	302.0	309.40	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End	
2	Sam Mendes	602.0	200.07	Action|Adventure|Thriller	Christoph Waltz	Spectre	
3	Christopher Nolan	813.0	448.13	Action|Thriller	Tom Hardy	The Dark Knight Rises	
5	Andrew Stanton	462.0	73.06	Action|Adventure|Sci-Fi	Daryl Sabara	John Carter	
...	...	...	...	...	...	...	
5033	Shane Carruth	143.0	0.42	Drama|Sci-Fi|Thriller	Shane Carruth	Primer	
5034	Neill Dela Llana	35.0	0.07	Thriller	Ian Gamazon	Cavite	
5035	Robert Rodriguez	56.0	2.04	Action|Crime|Drama|Romance|Thriller	Carlos Gallardo	El Mariachi	
5037	Edward Burns	14.0	0.00	Comedy|Drama	Kerry Bishé	Newlyweds	
5042	Jon Gunn	43.0	0.09	Documentary	John August	My Date with Drew	

3891 rows × 14 columns

In [154…

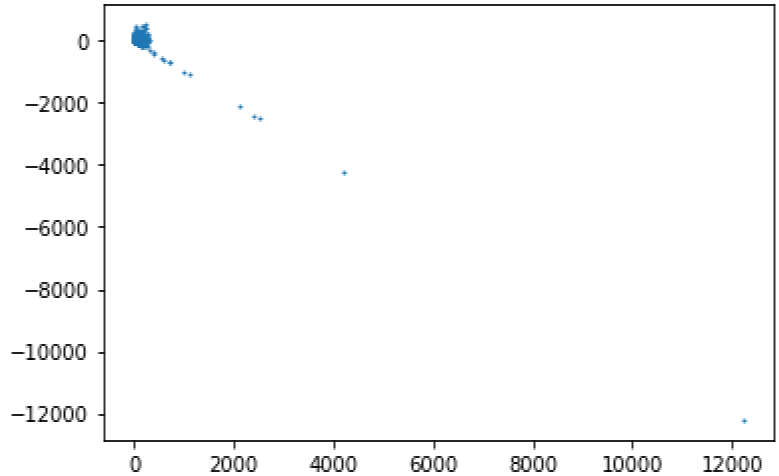
```
print(actual_data.sort_values('profit', ascending = False))
```

```
         director_name  num_critic_for_reviews   gross  \
0        James Cameron                   723.0  760.51
29     Colin Trevorrow                   644.0  652.18
26       James Cameron                   315.0  658.67
3024      George Lucas                   282.0  460.94
3080  Steven Spielberg                   215.0  434.95
...                ...                     ...     ...
2334   Katsuhiro Ôtomo                   105.0    0.41
2323    Hayao Miyazaki                   174.0    2.30
3005      Lajos Koltai                    73.0    0.20
3859    Chan-wook Park                   202.0    0.21
2988      Joon-ho Bong                   363.0    2.20

                                  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    English    237.00      2009.0         7.9
29                   1290    English    150.00      2015.0         7.0
26                   2528    English    200.00      1997.0         7.7
3024                 1470    English     11.00      1977.0         8.7
3080                  515    English     10.50      1982.0         7.9
...                   ...        ...       ...         ...         ...
2334                   79   Japanese   2127.52      2004.0         6.9
2323                  570   Japanese   2400.00      1997.0         8.4
3005                   45  Hungarian   2500.00      2005.0         7.1
3859                  131     Korean   4200.00      2005.0         7.7
2988                  279     Korean  12215.50      2006.0         7.0

      movie_facebook_likes    profit
0                    33000    523.51
29                  150000    502.18
26                   26000    458.67
3024                 33000    449.94
3080                 34000    424.45
...                    ...       ...
2334                   973  -2127.11
2323                 11000  -2397.70
3005                   607  -2499.80
3859                  4000  -4199.79
2988                  7000 -12213.30

[3891 rows x 14 columns]
```

In [155…

```
plt.scatter(actual_data.budget, actual_data.profit, s=1)
plt.figure(figsize=(20,20))
plt.show()
```




```
<Figure size 1440x1440 with 0 Axes>
```

In [156…

```
top10 = actual_data.sort_values('profit', ascending=False).head(10)
print(top10)

# We used Python library functions to find out the profit and the movies with the highest
```

```
          director_name  num_critic_for_reviews   gross  \
0         James Cameron                   723.0  760.51
29      Colin Trevorrow                   644.0  652.18
26        James Cameron                   315.0  658.67
3024       George Lucas                   282.0  460.94
3080   Steven Spielberg                   215.0  434.95
794         Joss Whedon                   703.0  623.28
17          Joss Whedon                   703.0  623.28
509        Roger Allers                   186.0  422.78
240        George Lucas                   320.0  474.54
66    Christopher Nolan                   645.0  533.32

                                 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  English   237.0      2009.0         7.9
29                   1290  English   150.0      2015.0         7.0
26                   2528  English   200.0      1997.0         7.7
3024                 1470  English    11.0      1977.0         8.7
3080                  515  English    10.5      1982.0         7.9
```

```
794                  1722  English   220.0      2012.0         8.1
17                   1722  English   220.0      2012.0         8.1
509                   656  English    45.0      1994.0         8.5
240                  3597  English   115.0      1999.0         6.5
66                   4667  English   185.0      2008.0         9.0

      movie_facebook_likes  profit
0                    33000  523.51
29                  150000  502.18
26                   26000  458.67
3024                 33000  449.94
3080                 34000  424.45
794                 123000  403.28
17                  123000  403.28
509                  17000  377.78
240                  13000  359.54
66                   37000  348.32
```

- ### 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`. Note that the same `movie_title` can be there in different languages.

In [161…

```
# Write your code for dropping duplicate values here
actual_data = actual_data.drop_duplicates()
actual_data
```

Out[161…

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title	nu
0	James Cameron	723.0	760.51	Action|Adventure|Fantasy|Sci-Fi	CCH Pounder	Avatar	
1	Gore Verbinski	302.0	309.40	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End	
2	Sam Mendes	602.0	200.07	Action|Adventure|Thriller	Christoph Waltz	Spectre	
3	Christopher Nolan	813.0	448.13	Action|Thriller	Tom Hardy	The Dark Knight Rises	
5	Andrew Stanton	462.0	73.06	Action|Adventure|Sci-Fi	Daryl Sabara	John Carter	
...	...	...	...	...	...	...	
5033	Shane Carruth	143.0	0.42	Drama|Sci-Fi|Thriller	Shane Carruth	Primer	
5034	Neill Dela Llana	35.0	0.07	Thriller	Ian Gamazon	Cavite	
5035	Robert Rodriguez	56.0	2.04	Action|Crime|Drama|Romance|Thriller	Carlos Gallardo	El Mariachi	
5037	Edward Burns	14.0	0.00	Comedy|Drama	Kerry Bishé	Newlyweds	
5042	Jon Gunn	43.0	0.09	Documentary	John August	My Date with Drew	

3856 rows × 14 columns

In [174…

```
# Write code for repeating subtask 2 here
top10 = actual_data.sort_values('profit', ascending=False).head(10).reset_index()
print(top10)


# Since we missed to filter some of the Duplicate Values in the DataSet,
# we had to remove them now and redo some of the tasks.
# Below is the Dataset displaying the  Top 10 movies based on the Profit Earning.
```

```
   index      director_name  num_critic_for_reviews   gross  \
0      0      James Cameron                   723.0  760.51
1     29    Colin Trevorrow                   644.0  652.18
2     26      James Cameron                   315.0  658.67
3   3024       George Lucas                   282.0  460.94
4   3080   Steven Spielberg                   215.0  434.95
5     17        Joss Whedon                   703.0  623.28
6    509       Roger Allers                   186.0  422.78
7    240       George Lucas                   320.0  474.54
8     66  Christopher Nolan                   645.0  533.32
9    439          Gary Ross                   673.0  408.00

                                genres         actor_1_name  \
0          Action|Adventure|Fantasy|Sci-Fi          CCH Pounder
1         Action|Adventure|Sci-Fi|Thriller  Bryce Dallas Howard
2                            Drama|Romance    Leonardo DiCaprio
3          Action|Adventure|Fantasy|Sci-Fi        Harrison Ford
4                            Family|Sci-Fi         Henry Thomas
5                  Action|Adventure|Sci-Fi      Chris Hemsworth
6  Adventure|Animation|Drama|Family|Musical    Matthew Broderick
7          Action|Adventure|Fantasy|Sci-Fi      Natalie Portman
8              Action|Crime|Drama|Thriller       Christian Bale
9          Adventure|Drama|Sci-Fi|Thriller    Jennifer Lawrence

                              movie_title  num_voted_users  \
0                                 Avatar            886204
1                         Jurassic World            418214
2                                Titanic            793059
3        Star Wars: Episode IV - A New Hope            911097
4                E.T. the Extra-Terrestrial            281842
5                           The Avengers            995415
6                          The Lion King            644348
7  Star Wars: Episode I - The Phantom Menace            534658
8                        The Dark Knight           1676169
9                       The Hunger Games            701607

  num_user_for_reviews language  budget  title_year  imdb_score  \
0                 3054  English   237.0      2009.0         7.9
1                 1290  English   150.0      2015.0         7.0
2                 2528  English   200.0      1997.0         7.7
3                 1470  English    11.0      1977.0         8.7
4                  515  English    10.5      1982.0         7.9
5                 1722  English   220.0      2012.0         8.1
6                  656  English    45.0      1994.0         8.5
7                 3597  English   115.0      1999.0         6.5
8                 4667  English   185.0      2008.0         9.0
9                 1959  English    78.0      2012.0         7.3

   movie_facebook_likes  profit
0                 33000  523.51
1                150000  502.18
2                 26000  458.67
3                 33000  449.94
4                 34000  424.45
```

```
5                123000  403.28
6                 17000  377.78
7                 13000  359.54
8                 37000  348.32
9                140000  330.00
```

Checkpoint 2: You might spot two movies directed by `James Cameron` in the list.

- ### 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` .

In [178…

```
# Write your code for extracting the top 250 movies as per the IMDb score here. Make sure
# and name that dataframe as 'IMDb_Top_250'

IMDb_Top_250 = actual_data[actual_data['num_voted_users'] > 25000].sort_values("imdb_score
IMDb_Top_250['Rank'] = [int(x) for x in IMDb_Top_250['imdb_score'].rank(ascending = False)
print(IMDb_Top_250)

# Created a Dataset displaying the Top 250 Movies based on there IMDb score and number of
```

```
            director_name  num_critic_for_reviews   gross  \
1937       Frank Darabont                   199.0   28.34
3466 Francis Ford Coppola                   208.0  134.82
2837 Francis Ford Coppola                   149.0   57.30
66      Christopher Nolan                   645.0  533.32
4498        Sergio Leone                   181.0    6.10
...                   ...                     ...     ...
4931         John Carney                   232.0    9.44
2605             Ang Lee                   287.0  128.07
3029     David O. Russell                   410.0   93.57
2177          Tim Burton                   111.0   56.36
2487        George Cukor                    82.0   72.00

                           genres     actor_1_name  \
1937                  Crime|Drama   Morgan Freeman
3466                  Crime|Drama        Al Pacino
2837                  Crime|Drama  Robert De Niro
66    Action|Crime|Drama|Thriller  Christian Bale
4498                      Western   Clint Eastwood
...                           ...              ...
4931          Drama|Music|Romance     Glen Hansard
2605        Action|Drama|Romance       Chen Chang
3029       Biography|Drama|Sport   Christian Bale
2177              Fantasy|Romance      Johnny Depp
2487 Drama|Family|Musical|Romance     Jeremy Brett

                          movie_title  num_voted_users num_user_for_reviews  \
1937         The Shawshank Redemption          1689764                 4144
3466                    The Godfather          1155770                 2238
2837           The Godfather: Part II           790926                  650
66                    The Dark Knight          1676169                 4667
4498   The Good, the Bad and the Ugly           503509                  780
...                               ...              ...                  ...
4931                             Once            90827                  329
2605  Crouching Tiger, Hidden Dragon           217740                 1641
3029                      The Fighter           275869                  389
```

```
2177              Edward Scissorhands             357581                588
2487                     My Fair Lady              66959                258

      language  budget  title_year  imdb_score  movie_facebook_likes  profit  \
1937   English   25.00      1994.0         9.3                108000    3.34
3466   English    6.00      1972.0         9.2                 43000  128.82
2837   English   13.00      1974.0         9.0                 14000   44.30
66     English  185.00      2008.0         9.0                 37000  348.32
4498   Italian    1.20      1966.0         8.9                 20000    4.90
...        ...     ...         ...         ...                   ...     ...
4931   English    0.18      2007.0         7.9                 26000    9.26
2605  Mandarin   15.00      2000.0         7.9                     0  113.07
3029   English   25.00      2010.0         7.9                 36000   68.57
2177   English   20.00      1990.0         7.9                 16000   36.36
2487   English   17.00      1964.0         7.9                     0   55.00

      Rank
1937     1
3466     2
2837     3
66       3
4498     6
...    ...
4931   228
2605   228
3029   228
2177   228
2487   228

[250 rows x 15 columns]
```

In [179…

```
# Write your code to extract top foreign language films from 'IMDb_Top_250' here

Top_Foreign_Lang_Film = IMDb_Top_250[IMDb_Top_250["language"] != "English"]
print(Top_Foreign_Lang_Film)

# Displayed the top Foreings Films.
```

```
                        director_name  num_critic_for_reviews   gross  \
4498                     Sergio Leone                   181.0    6.10
4747                   Akira Kurosawa                   153.0    0.27
4029               Fernando Meirelles                   214.0    7.56
2373                   Hayao Miyazaki                   246.0   10.05
4259  Florian Henckel von Donnersmarck                   215.0   11.28
4921                     Majid Majidi                    46.0    0.93
2323                   Hayao Miyazaki                   174.0    2.30
2970                Wolfgang Petersen                    96.0   11.43
4105                   Chan-wook Park                   305.0    2.18
4659                   Asghar Farhadi                   354.0    7.10
1329                   S.S. Rajamouli                    44.0    6.50
1298               Jean-Pierre Jeunet                   242.0   33.20
2734                       Fritz Lang                   260.0    0.03
4033                Thomas Vinterberg                   349.0    0.61
2829              Oliver Hirschbiegel                   192.0    5.50
2551               Guillermo del Toro                   406.0   37.62
4000              Juan José Campanella                   262.0   20.17
3550                  Denis Villeneuve                   226.0    6.86
2047                   Hayao Miyazaki                   212.0    4.71
2830               Alejandro Amenábar                   157.0    2.09
2914                      Je-kyu Kang                    86.0    1.11
4461                Thomas Vinterberg                    98.0    1.65
3553                     José Padilha                   142.0    0.01
3423                  Katsuhiro Ôtomo                   150.0    0.44
4267              Alejandro G. Iñárritu                   157.0    5.38
3456                Vincent Paronnaud                   242.0    4.44
```

```
3344                Karan Johar                    210.0    4.02
4144              Walter Salles                     71.0    5.60
4284                 Ari Folman                    231.0    2.28
4897              Sergio Leone                     122.0    3.50
1171                Yimou Zhang                    283.0    0.08
2863             Clint Eastwood                    251.0   13.75
3264             Michael Haneke                    447.0    0.23
3510                Yash Chopra                     29.0    2.92
3677       Christophe Barratier                    112.0    3.63
4415           Fabián Bielinsky                     94.0    1.22
4640            Cristian Mungiu                    233.0    1.19
2605                    Ang Lee                    287.0  128.07

                                              genres  \
4498                                         Western
4747                          Action|Adventure|Drama
4029                                     Crime|Drama
2373               Adventure|Animation|Family|Fantasy
4259                                  Drama|Thriller
4921                                    Drama|Family
2323                      Adventure|Animation|Fantasy
2970                       Adventure|Drama|Thriller|War
4105                           Drama|Mystery|Thriller
4659                                   Drama|Mystery
1329               Action|Adventure|Drama|Fantasy|War
1298                                  Comedy|Romance
2734                                    Drama|Sci-Fi
4033                                           Drama
2829                      Biography|Drama|History|War
2551                               Drama|Fantasy|War
4000                           Drama|Mystery|Thriller
3550                               Drama|Mystery|War
2047               Adventure|Animation|Family|Fantasy
2830                         Biography|Drama|Romance
2914                                Action|Drama|War
4461                                           Drama
3553                      Action|Crime|Drama|Thriller
3423                         Action|Animation|Sci-Fi
4267                                  Drama|Thriller
3456                     Animation|Biography|Drama|War
3344                        Adventure|Drama|Thriller
4144                                           Drama
4284  Animation|Biography|Documentary|Drama|History|War
4897                           Action|Drama|Western
1171                        Action|Adventure|History
2863                               Drama|History|War
3264                                   Drama|Romance
3510                           Drama|Musical|Romance
3677                                     Drama|Music
4415                           Crime|Drama|Thriller
4640                                           Drama
2605                           Action|Drama|Romance

                actor_1_name                          movie_title  \
4498          Clint Eastwood      The Good, the Bad and the Ugly
4747         Takashi Shimura                       Seven Samurai
4029             Alice Braga                         City of God
2373          Bunta Sugawara                      Spirited Away
4259         Sebastian Koch                 The Lives of Others
4921          Bahare Seddiqi                  Children of Heaven
2323           Minnie Driver                   Princess Mononoke
2970         Jürgen Prochnow                            Das Boot
4105            Min-sik Choi                              Oldboy
4659         Shahab Hosseini                        A Separation
1329        Tamannaah Bhatia             Baahubali: The Beginning
1298       Mathieu Kassovitz                               Amélie
```

2734	Brigitte Helm	Metropolis
4033	Thomas Bo Larsen	The Hunt
2829	Thomas Kretschmann	Downfall
2551	Ivana Baquero	Pan's Labyrinth
4000	Ricardo Darín	The Secret in Their Eyes
3550	Lubna Azabal	Incendies
2047	Christian Bale	Howl's Moving Castle
2830	Belén Rueda	The Sea Inside
2914	Min-sik Choi	Tae Guk Gi: The Brotherhood of War
4461	Ulrich Thomsen	The Celebration
3553	Wagner Moura	Elite Squad
3423	Mitsuo Iwata	Akira
4267	Adriana Barraza	Amores Perros
3456	Catherine Deneuve	Persepolis
3344	Shah Rukh Khan	My Name Is Khan
4144	Fernanda Montenegro	Central Station
4284	Ari Folman	Waltz with Bashir
4897	Clint Eastwood	A Fistful of Dollars
1171	Jet Li	Hero
2863	Yuki Matsuzaki	Letters from Iwo Jima
3264	Isabelle Huppert	Amour
3510	Shah Rukh Khan	Veer-Zaara
3677	Jean-Baptiste Maunier	The Chorus
4415	Ricardo Darín	Nine Queens
4640	Anamaria Marinca	4 Months, 3 Weeks and 2 Days
2605	Chen Chang	Crouching Tiger, Hidden Dragon

	num_voted_users	num_user_for_reviews	language	budget	title_year \
4498	503509	780	Italian	1.20	1966.0
4747	229012	596	Japanese	2.00	1954.0
4029	533200	749	Portuguese	3.30	2002.0
2373	417971	902	Japanese	19.00	2001.0
4259	259379	407	German	2.00	2006.0
4921	27882	130	Persian	0.18	1997.0
2323	221552	570	Japanese	2400.00	1997.0
2970	168203	426	German	14.00	1981.0
4105	356181	809	Korean	3.00	2003.0
4659	151812	264	Persian	0.50	2011.0
1329	62756	410	Telugu	18.03	2015.0
1298	534262	1314	French	77.00	2001.0
2734	111841	413	German	6.00	1927.0
4033	170155	249	Danish	3.80	2012.0
2829	248354	564	German	13.50	2004.0
2551	467234	1083	Spanish	13.50	2006.0
4000	131831	231	Spanish	2.00	2009.0
3550	80429	156	French	6.80	2010.0
2047	214091	330	Japanese	24.00	2004.0
2830	64556	140	Spanish	10.00	2004.0
2914	31943	224	Korean	12.80	2004.0
4461	65951	258	Danish	1.30	1998.0
3553	81644	107	Portuguese	4.00	2007.0
3423	106160	430	Japanese	1100.00	1988.0
4267	173551	361	Spanish	2.00	2000.0
3456	70194	158	French	7.30	2007.0
3344	69759	235	Hindi	12.00	2010.0
4144	28951	257	Portuguese	2.90	1998.0
4284	46107	156	Hebrew	1.50	2008.0
4897	147566	235	Italian	0.20	1964.0
1171	149414	841	Mandarin	31.00	2002.0
2863	132149	316	Japanese	19.00	2006.0
3264	70382	190	French	8.90	2012.0
3510	34449	119	Hindi	7.00	2004.0
3677	44151	110	French	5.50	2004.0
4415	38215	125	Spanish	1.50	2000.0
4640	44763	172	Romanian	0.59	2007.0
2605	217740	1641	Mandarin	15.00	2000.0

```
      imdb_score  movie_facebook_likes    profit  Rank
4498         8.9                 20000      4.90     6
4747         8.7                 11000     -1.73    17
4029         8.7                 28000      4.26    17
2373         8.6                 28000     -8.95    24
4259         8.5                 39000      9.28    37
4921         8.5                     0      0.75    37
2323         8.4                 11000  -2397.70    54
2970         8.4                 11000     -2.57    54
4105         8.4                 43000     -0.82    54
4659         8.4                 48000      6.60    54
1329         8.4                 21000    -11.53    54
1298         8.4                 39000    -43.80    54
2734         8.3                 12000     -5.97    73
4033         8.3                 60000     -3.19    73
2829         8.3                 14000     -8.00    73
2551         8.2                 27000     24.12    96
4000         8.2                 33000     18.17    96
3550         8.2                 37000      0.06    96
2047         8.2                 13000    -19.29    96
2830         8.1                     0     -7.91   130
2914         8.1                     0    -11.69   130
4461         8.1                  5000      0.35   130
3553         8.1                 11000     -3.99   130
3423         8.1                     0  -1099.56   130
4267         8.1                 11000      3.38   130
3456         8.0                 14000     -2.86   179
3344         8.0                 27000     -7.98   179
4144         8.0                     0      2.70   179
4284         8.0                     0      0.78   179
4897         8.0                     0      3.30   179
1171         7.9                     0    -30.92   228
2863         7.9                  5000     -5.25   228
3264         7.9                 33000     -8.67   228
3510         7.9                  2000     -4.08   228
3677         7.9                     0     -1.87   228
4415         7.9                     0     -0.28   228
4640         7.9                 14000      0.60   228
2605         7.9                     0    113.07   228
```

Checkpoint 3: Can you spot `Veer-Zaara` in the dataframe?

- ### 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` . Incase of a tie in IMDb score between two directors, sort them alphabetically.

In [183…

```
# Write your code for extracting the top 10 directors here

top10director = actual_data.groupby('director_name')['imdb_score'].mean().sort_values(asce
print(top10director)

# We made a list of top 10 best Directors of all time and Charles Chaplin tops the list.
```

```
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
```

Checkpoint 4: No surprises that `Damien Chazelle` (director of Whiplash and La La Land) is in this list.

- ### 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` will 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`.

In [189…

```
# Write your code for extracting the first two genres of each movie here

actual_data[['genre_1', 'genre_2', 'remaining']] = actual_data['genres'].str.split('|', n=
actual_data.drop('remaining', inplace=True, axis=1)
actual_data["genre_2"] = actual_data["genre_2"].fillna(actual_data["genre_1"])
print(actual_data)

# Since we had Movie Genre given in a Single Column, we had to split the Genre into Genre
```

```
           director_name  num_critic_for_reviews   gross  \
0          James Cameron                   723.0  760.51
1         Gore Verbinski                   302.0  309.40
2             Sam Mendes                   602.0  200.07
3      Christopher Nolan                   813.0  448.13
5         Andrew Stanton                   462.0   73.06
...                  ...                     ...     ...
5033       Shane Carruth                   143.0    0.42
5034     Neill Dela Llana                   35.0    0.07
5035     Robert Rodriguez                   56.0    2.04
5037        Edward Burns                    14.0    0.00
5042            Jon Gunn                    43.0    0.09

                                  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  English  237.00      2009.0         7.9
1                    1238  English  300.00      2007.0         7.1
2                     994  English  245.00      2015.0         6.8
3                    2701  English  250.00      2012.0         8.5
5                     738  English  263.70      2012.0         6.6
...                   ...      ...     ...         ...         ...
5033                  371  English    0.01      2004.0         7.0
5034                   35  English    0.01      2005.0         6.3
5035                  130  Spanish    0.01      1992.0         6.9
5037                   14  English    0.01      2011.0         6.4
5042                   84  English    0.00      2004.0         6.6

      movie_facebook_likes  profit      genre_1      genre_2
0                    33000  523.51       Action    Adventure
1                        0    9.40       Action    Adventure
2                    85000  -44.93       Action    Adventure
3                   164000  198.13       Action     Thriller
5                    24000 -190.64       Action    Adventure
...                    ...     ...          ...          ...
5033                 19000    0.41        Drama       Sci-Fi
5034                    74    0.06     Thriller     Thriller
5035                     0    2.03       Action        Crime
5037                   413   -0.01       Comedy        Drama
5042                   456    0.09  Documentary  Documentary

[3856 rows x 16 columns]
```

In [191…

```python
# Write your code for grouping the dataframe here

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

In [193…

```python
# Write your code for getting the 5 most popular combo of genres here

PopGenre = movies_by_segment.gross.mean()
PopGenre = PopGenre.sort_values(ascending = False)
print(PopGenre)

# Here we were able to find out the best combination of the two genres and it was  Family
```

```
genre_1    genre_2
Family     Sci-Fi       434.950000
Adventure  Sci-Fi       228.628750
           Family       118.918824
           Animation    116.998462
Action     Adventure    109.595510
                           ...
Horror     Musical        0.140000
Romance    Romance        0.105000
Thriller   Thriller       0.040000
Sci-Fi     Sci-Fi         0.020000
Adventure  War            0.010000
Name: gross, Length: 103, dtype: float64
```

Checkpoint 5: Well, as it turns out. `Family + Sci-Fi` is the most popular combo of genres out there!

- ### 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_users_for_review` and identify the actors which have the highest mean.
    5. Observe the change in number of voted users over decades using a bar chart. Create a column called `decade` which represents the decade to which every movie belongs to. For example, the `title_year` year 1923, 1925 should be stored as 1920s. Sort the dataframe based on the column `decade`, group it by `decade` and find the sum of users voted in each decade. Store this in a new data frame called `df_by_decade`.

In [226…

```python
# Write your code for creating three new dataframes here

Meryl_Streep = actual_data[actual_data['actor_1_name'] == "Meryl Streep"]
print(Meryl_Streep)

# Include all movies in which Meryl_Streep is the lead
```

```
        director_name  num_critic_for_reviews   gross  \
410      Nancy Meyers                 187.000 112.700
1106    Curtis Hanson                  42.000  46.820
1204      Nora Ephron                 252.000  94.130
1408    David Frankel                 208.000 124.730
1483   Robert Redford                 227.000  15.000
1575   Sydney Pollack                  66.000  87.100
1618    David Frankel                 234.000  63.540
1674    Carl Franklin                  64.000  23.210
1925   Stephen Daldry                 174.000  41.600
2781   Phyllida Lloyd                 331.000  29.960
3135    Robert Altman                 211.000  20.340

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

                    movie_title  num_voted_users num_user_for_reviews  \
410            It's Complicated            69860                  214
1106             The River Wild            32544                   69
1204             Julie & Julia            79264                  277
1408       The Devil Wears Prada           286178                  631
1483            Lions for Lambs            41170                  298
1575              Out of Africa            52339                  200
1618               Hope Springs            34258                  178
1674             One True Thing             9283                  112
1925                  The Hours           102123                  660
2781              The Iron Lady            82327                  350
```

```
3135  A Prairie Home Companion             19655              280

    language  budget  title_year  imdb_score  movie_facebook_likes  profit   \
410   English  85.000    2009.000       6.600                     0  27.700
1106  English  45.000    1994.000       6.300                     0   1.820
1204  English  40.000    2009.000       7.000                 13000  54.130
1408  English  35.000    2006.000       6.800                     0  89.730
1483  English  35.000    2007.000       6.200                     0 -20.000
1575  English  31.000    1985.000       7.200                     0  56.100
1618  English  30.000    2012.000       6.300                     0  33.540
1674  English  30.000    1998.000       7.000                   592  -6.790
1925  English  25.000    2002.000       7.600                     0  16.600
2781  English  13.000    2011.000       6.400                 18000  16.960
3135  English  10.000    2006.000       6.800                   683  10.340

        genre_1    genre_2
410      Comedy      Drama
1106     Action  Adventure
1204  Biography      Drama
1408     Comedy      Drama
1483      Drama   Thriller
1575  Biography      Drama
1618     Comedy      Drama
1674      Drama      Drama
1925      Drama    Romance
2781  Biography      Drama
3135     Comedy      Drama
```

In [227…

```python
Leo_Caprio = actual_data[actual_data['actor_1_name'] == "Leonardo DiCaprio"]
print(LeoCaprio)
```

```
               director_name  num_critic_for_reviews   gross  \
26             James Cameron                 315.000 658.670
50              Baz Luhrmann                 490.000 144.810
97         Christopher Nolan                 642.000 292.570
179    Alejandro G. Iñárritu                 556.000 183.640
257          Martin Scorsese                 267.000 102.610
296        Quentin Tarantino                 765.000 162.800
307             Edward Zwick                 166.000  57.370
308          Martin Scorsese                 606.000 116.870
326          Martin Scorsese                 233.000  77.680
361          Martin Scorsese                 352.000 132.370
452          Martin Scorsese                 490.000 127.970
641             Ridley Scott                 238.000  39.380
911         Steven Spielberg                 194.000 164.440
990              Danny Boyle                 118.000  39.780
1114              Sam Mendes                 323.000  22.880
1422          Randall Wallace                 83.000  56.880
1453          Clint Eastwood                 392.000  37.300
1560               Sam Raimi                  63.000  18.640
2067              Jerry Zaks                  45.000  12.780
2757            Baz Luhrmann                 106.000  46.340
3476            Baz Luhrmann                 490.000 144.810

                               genres       actor_1_name  \
26                      Drama|Romance  Leonardo DiCaprio
50                      Drama|Romance  Leonardo DiCaprio
97    Action|Adventure|Sci-Fi|Thriller  Leonardo DiCaprio
179   Adventure|Drama|Thriller|Western  Leonardo DiCaprio
257                   Biography|Drama  Leonardo DiCaprio
296                     Drama|Western  Leonardo DiCaprio
307          Adventure|Drama|Thriller  Leonardo DiCaprio
308      Biography|Comedy|Crime|Drama  Leonardo DiCaprio
326                       Crime|Drama  Leonardo DiCaprio
361              Crime|Drama|Thriller  Leonardo DiCaprio
```

```
452                 Mystery|Thriller  Leonardo DiCaprio
641            Action|Drama|Thriller  Leonardo DiCaprio
911             Biography|Crime|Drama  Leonardo DiCaprio
990         Adventure|Drama|Thriller  Leonardo DiCaprio
1114                   Drama|Romance  Leonardo DiCaprio
1422                Action|Adventure  Leonardo DiCaprio
1453           Biography|Crime|Drama  Leonardo DiCaprio
1560          Action|Thriller|Western  Leonardo DiCaprio
2067                           Drama  Leonardo DiCaprio
2757                   Drama|Romance  Leonardo DiCaprio
3476                   Drama|Romance  Leonardo DiCaprio

                   movie_title  num_voted_users num_user_for_reviews  \
26                     Titanic           793059                 2528
50            The Great Gatsby           362912                  753
97                   Inception          1468200                 2803
179                The Revenant           406020                 1188
257                 The Aviator           264318                  799
296           Django Unchained           955174                 1193
307              Blood Diamond           400292                  657
308    The Wolf of Wall Street           780588                 1138
326          Gangs of New York           314033                 1166
361               The Departed           873649                 2054
452              Shutter Island           786092                  964
641               Body of Lies           174248                  263
911         Catch Me If You Can           525801                  667
990                  The Beach           176169                  548
1114         Revolutionary Road           152591                  414
1422  The Man in the Iron Mask           125219                  244
1453                  J. Edgar           102728                  279
1560     The Quick and the Dead            69197                  216
2067              Marvin's Room            20163                   71
2757             Romeo + Juliet           167750                  506
3476           The Great Gatsby           362933                  753

     language  budget  title_year  imdb_score  movie_facebook_likes  profit  \
26    English 200.000    1997.000       7.700                 26000 458.670
50    English 105.000    2013.000       7.300                115000  39.810
97    English 160.000    2010.000       8.800                175000 132.570
179   English 135.000    2015.000       8.100                190000  48.640
257   English 110.000    2004.000       7.500                     0  -7.390
296   English 100.000    2012.000       8.500                199000  62.800
307   English 100.000    2006.000       8.000                 14000 -42.630
308   English 100.000    2013.000       8.200                138000  16.870
326   English 100.000    2002.000       7.500                     0 -22.320
361   English  90.000    2006.000       8.500                 29000  42.370
452   English  80.000    2010.000       8.100                 53000  47.970
641   English  70.000    2008.000       7.100                     0 -30.620
911   English  52.000    2002.000       8.000                 15000 112.440
990   English  50.000    2000.000       6.600                     0 -10.220
1114  English  35.000    2008.000       7.300                     0 -12.120
1422  English  35.000    1998.000       6.400                     0  21.880
1453  English  35.000    2011.000       6.600                 16000   2.300
1560  English  32.000    1995.000       6.400                     0 -13.360
2067  English  23.000    1996.000       6.700                  1000 -10.220
2757  English  14.500    1996.000       6.800                 10000  31.840
3476  English 105.000    2013.000       7.300                115000  39.810

        genre_1    genre_2
26        Drama    Romance
50        Drama    Romance
97       Action  Adventure
179   Adventure      Drama
257   Biography      Drama
296       Drama    Western
307   Adventure      Drama
```

```
308   Biography     Comedy
326       Crime      Drama
361       Crime      Drama
452     Mystery   Thriller
641      Action      Drama
911   Biography      Crime
990   Adventure      Drama
1114      Drama    Romance
1422     Action  Adventure
1453  Biography      Crime
1560     Action   Thriller
2067      Drama      Drama
2757      Drama    Romance
3476      Drama    Romance
```

In [224…

```python
Brad_Pitt = actual_data[actual_data['actor_1_name'] == "Brad Pitt"]
print(Brad_Pitt)
```

```
             director_name  num_critic_for_reviews   gross  \
101          David Fincher                 362.000 127.490
147      Wolfgang Petersen                 220.000 133.230
254     Steven Soderbergh                 198.000 125.530
255             Doug Liman                 233.000 186.340
382             Tony Scott                 142.000   0.030
400     Steven Soderbergh                 186.000 183.410
470             David Ayer                 406.000  85.710
611    Jean-Jacques Annaud                  76.000  37.900
683          David Fincher                 315.000  37.020
792        Patrick Gilmore                  98.000  26.290
940            Neil Jordan                 120.000 105.260
1490       Terrence Malick                 584.000  13.300
1722        Andrew Dominik                 273.000   3.900
2204  Alejandro G. Iñárritu                285.000  34.300
2333    Angelina Jolie Pitt                131.000   0.530
2682        Andrew Dominik                 414.000  14.940
2898            Tony Scott                 122.000  12.280

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

                                     movie_title  num_voted_users  \
101           The Curious Case of Benjamin Button           459346
147                                          Troy           381672
254                                Ocean's Twelve           284852
255                              Mr. & Mrs. Smith           348861
382                                       Spy Game           121259
400                                Ocean's Eleven           402645
470                                           Fury           303185
611                           Seven Years in Tibet            96385
```

```
683                                             Fight Club           1347461
792                         Sinbad: Legend of the Seven Seas           36144
940   Interview with the Vampire: The Vampire Chroni...            239752
1490                                       The Tree of Life          136367
1722  The Assassination of Jesse James by the Coward...            136104
2204                                                  Babel          243799
2333                                             By the Sea            7976
2682                                   Killing Them Softly          111625
2898                                          True Romance          163492

     num_user_for_reviews language  budget  title_year  imdb_score  \
101                   822  English 150.000    2008.000       7.800   
147                  1694  English 175.000    2004.000       7.200   
254                   627  English 110.000    2004.000       6.400   
255                   798  English 120.000    2005.000       6.500   
382                   361  English  92.000    2001.000       7.000   
400                   845  English  85.000    2001.000       7.800   
470                   701  English  68.000    2014.000       7.600   
611                   119  English  70.000    1997.000       7.000   
683                  2968  English  63.000    1999.000       8.800   
792                    91  English  60.000    2003.000       6.700   
940                   406  English  60.000    1994.000       7.600   
1490                  975  English  32.000    2011.000       6.700   
1722                  415  English  30.000    2007.000       7.500   
2204                  908  English  25.000    2006.000       7.500   
2333                   61  English  10.000    2015.000       5.300   
2682                  369  English  15.000    2012.000       6.200   
2898                  460  English  13.000    1993.000       8.000   

      movie_facebook_likes  profit    genre_1    genre_2  
101                  23000 -22.510      Drama    Fantasy  
147                      0 -41.770  Adventure  Adventure  
254                      0  15.530      Crime   Thriller  
255                      0  66.340     Action     Comedy  
382                      0 -91.970     Action      Crime  
400                      0  98.410      Crime   Thriller  
470                  82000  17.710     Action      Drama  
611                      0 -32.100  Adventure  Biography  
683                  48000 -25.980      Drama      Drama  
792                    880 -33.710  Adventure  Animation  
940                  11000  45.260      Drama    Fantasy  
1490                 39000 -18.700      Drama    Fantasy  
1722                     0 -26.100  Biography      Crime  
2204                     0   9.300      Drama      Drama  
2333                     0  -9.470      Drama    Romance  
2682                 20000  -0.060      Crime   Thriller  
2898                 15000  -0.720     Action      Crime  
```

In [228…

```
# Write your code for combining the three dataframes here
Combined = pd.concat([Meryl_Streep, Leo_Caprio, Brad_Pitt], axis = 0)
print(Combined)

# We Created a Dataset only with movies that had Meryl_Streep, Leo_Caprio, Brad_Pitt as th
```

```
             director_name  num_critic_for_reviews   gross  \
410           Nancy Meyers                 187.000 112.700   
1106         Curtis Hanson                  42.000  46.820   
1204           Nora Ephron                 252.000  94.130   
1408         David Frankel                 208.000 124.730   
1483        Robert Redford                 227.000  15.000   
1575        Sydney Pollack                  66.000  87.100   
1618         David Frankel                 234.000  63.540   
1674          Carl Franklin                  64.000  23.210   
1925        Stephen Daldry                 174.000  41.600   
2781        Phyllida Lloyd                 331.000  29.960   
```

3135	Robert Altman	211.000	20.340
26	James Cameron	315.000	658.670
50	Baz Luhrmann	490.000	144.810
97	Christopher Nolan	642.000	292.570
179	Alejandro G. Iñárritu	556.000	183.640
257	Martin Scorsese	267.000	102.610
296	Quentin Tarantino	765.000	162.800
307	Edward Zwick	166.000	57.370
308	Martin Scorsese	606.000	116.870
326	Martin Scorsese	233.000	77.680
361	Martin Scorsese	352.000	132.370
452	Martin Scorsese	490.000	127.970
641	Ridley Scott	238.000	39.380
911	Steven Spielberg	194.000	164.440
990	Danny Boyle	118.000	39.780
1114	Sam Mendes	323.000	22.880
1422	Randall Wallace	83.000	56.880
1453	Clint Eastwood	392.000	37.300
1560	Sam Raimi	63.000	18.640
2067	Jerry Zaks	45.000	12.780
2757	Baz Luhrmann	106.000	46.340
3476	Baz Luhrmann	490.000	144.810
101	David Fincher	362.000	127.490
147	Wolfgang Petersen	220.000	133.230
254	Steven Soderbergh	198.000	125.530
255	Doug Liman	233.000	186.340
382	Tony Scott	142.000	0.030
400	Steven Soderbergh	186.000	183.410
470	David Ayer	406.000	85.710
611	Jean-Jacques Annaud	76.000	37.900
683	David Fincher	315.000	37.020
792	Patrick Gilmore	98.000	26.290
940	Neil Jordan	120.000	105.260
1490	Terrence Malick	584.000	13.300
1722	Andrew Dominik	273.000	3.900
2204	Alejandro G. Iñárritu	285.000	34.300
2333	Angelina Jolie Pitt	131.000	0.530
2682	Andrew Dominik	414.000	14.940
2898	Tony Scott	122.000	12.280

	genres	actor_1_name \
410	Comedy|Drama|Romance	Meryl Streep
1106	Action|Adventure|Crime|Thriller	Meryl Streep
1204	Biography|Drama|Romance	Meryl Streep
1408	Comedy|Drama|Romance	Meryl Streep
1483	Drama|Thriller|War	Meryl Streep
1575	Biography|Drama|Romance	Meryl Streep
1618	Comedy|Drama|Romance	Meryl Streep
1674	Drama	Meryl Streep
1925	Drama|Romance	Meryl Streep
2781	Biography|Drama|History	Meryl Streep
3135	Comedy|Drama|Music	Meryl Streep
26	Drama|Romance	Leonardo DiCaprio
50	Drama|Romance	Leonardo DiCaprio
97	Action|Adventure|Sci-Fi|Thriller	Leonardo DiCaprio
179	Adventure|Drama|Thriller|Western	Leonardo DiCaprio
257	Biography|Drama	Leonardo DiCaprio
296	Drama|Western	Leonardo DiCaprio
307	Adventure|Drama|Thriller	Leonardo DiCaprio
308	Biography|Comedy|Crime|Drama	Leonardo DiCaprio
326	Crime|Drama	Leonardo DiCaprio
361	Crime|Drama|Thriller	Leonardo DiCaprio
452	Mystery|Thriller	Leonardo DiCaprio
641	Action|Drama|Thriller	Leonardo DiCaprio
911	Biography|Crime|Drama	Leonardo DiCaprio
990	Adventure|Drama|Thriller	Leonardo DiCaprio

```
1114                                 Drama|Romance  Leonardo DiCaprio
1422                              Action|Adventure  Leonardo DiCaprio
1453                         Biography|Crime|Drama  Leonardo DiCaprio
1560                       Action|Thriller|Western  Leonardo DiCaprio
2067                                         Drama  Leonardo DiCaprio
2757                                 Drama|Romance  Leonardo DiCaprio
3476                                 Drama|Romance  Leonardo DiCaprio
101                          Drama|Fantasy|Romance          Brad Pitt
147                                      Adventure          Brad Pitt
254                                 Crime|Thriller          Brad Pitt
255             Action|Comedy|Crime|Romance|Thriller          Brad Pitt
382                          Action|Crime|Thriller          Brad Pitt
400                                 Crime|Thriller          Brad Pitt
470                               Action|Drama|War          Brad Pitt
611            Adventure|Biography|Drama|History|War          Brad Pitt
683                                          Drama          Brad Pitt
792   Adventure|Animation|Comedy|Drama|Family|Fantas...          Brad Pitt
940                            Drama|Fantasy|Horror          Brad Pitt
1490                                 Drama|Fantasy          Brad Pitt
1722           Biography|Crime|Drama|History|Western          Brad Pitt
2204                                         Drama          Brad Pitt
2333                                 Drama|Romance          Brad Pitt
2682                                Crime|Thriller          Brad Pitt
2898             Action|Crime|Drama|Romance|Thriller          Brad Pitt

                                       movie_title  num_voted_users  \
410                             It's Complicated            69860
1106                              The River Wild            32544
1204                              Julie & Julia            79264
1408                         The Devil Wears Prada           286178
1483                              Lions for Lambs            41170
1575                               Out of Africa            52339
1618                                Hope Springs            34258
1674                              One True Thing             9283
1925                                   The Hours           102123
2781                               The Iron Lady            82327
3135                      A Prairie Home Companion            19655
26                                       Titanic           793059
50                              The Great Gatsby           362912
97                                     Inception          1468200
179                                 The Revenant           406020
257                                  The Aviator           264318
296                             Django Unchained           955174
307                               Blood Diamond           400292
308                        The Wolf of Wall Street           780588
326                           Gangs of New York           314033
361                                 The Departed           873649
452                               Shutter Island           786092
641                                 Body of Lies           174248
911                            Catch Me If You Can           525801
990                                    The Beach           176169
1114                           Revolutionary Road           152591
1422                      The Man in the Iron Mask           125219
1453                                    J. Edgar           102728
1560                       The Quick and the Dead            69197
2067                               Marvin's Room            20163
2757                              Romeo + Juliet           167750
3476                            The Great Gatsby           362933
101              The Curious Case of Benjamin Button           459346
147                                         Troy           381672
254                               Ocean's Twelve           284852
255                              Mr. & Mrs. Smith           348861
382                                     Spy Game           121259
400                               Ocean's Eleven           402645
470                                         Fury           303185
611                          Seven Years in Tibet            96385
```

```
683                                           Fight Club          1347461
792                     Sinbad: Legend of the Seven Seas            36144
940   Interview with the Vampire: The Vampire Chroni...           239752
1490                                   The Tree of Life           136367
1722  The Assassination of Jesse James by the Coward...           136104
2204                                              Babel           243799
2333                                         By the Sea             7976
2682                                 Killing Them Softly          111625
2898                                       True Romance           163492

     num_user_for_reviews language  budget  title_year  imdb_score  \
410                   214  English  85.000    2009.000       6.600
1106                   69  English  45.000    1994.000       6.300
1204                  277  English  40.000    2009.000       7.000
1408                  631  English  35.000    2006.000       6.800
1483                  298  English  35.000    2007.000       6.200
1575                  200  English  31.000    1985.000       7.200
1618                  178  English  30.000    2012.000       6.300
1674                  112  English  30.000    1998.000       7.000
1925                  660  English  25.000    2002.000       7.600
2781                  350  English  13.000    2011.000       6.400
3135                  280  English  10.000    2006.000       6.800
26                   2528  English 200.000    1997.000       7.700
50                    753  English 105.000    2013.000       7.300
97                   2803  English 160.000    2010.000       8.800
179                  1188  English 135.000    2015.000       8.100
257                   799  English 110.000    2004.000       7.500
296                  1193  English 100.000    2012.000       8.500
307                   657  English 100.000    2006.000       8.000
308                  1138  English 100.000    2013.000       8.200
326                  1166  English 100.000    2002.000       7.500
361                  2054  English  90.000    2006.000       8.500
452                   964  English  80.000    2010.000       8.100
641                   263  English  70.000    2008.000       7.100
911                   667  English  52.000    2002.000       8.000
990                   548  English  50.000    2000.000       6.600
1114                  414  English  35.000    2008.000       7.300
1422                  244  English  35.000    1998.000       6.400
1453                  279  English  35.000    2011.000       6.600
1560                  216  English  32.000    1995.000       6.400
2067                   71  English  23.000    1996.000       6.700
2757                  506  English  14.500    1996.000       6.800
3476                  753  English 105.000    2013.000       7.300
101                   822  English 150.000    2008.000       7.800
147                  1694  English 175.000    2004.000       7.200
254                   627  English 110.000    2004.000       6.400
255                   798  English 120.000    2005.000       6.500
382                   361  English  92.000    2001.000       7.000
400                   845  English  85.000    2001.000       7.800
470                   701  English  68.000    2014.000       7.600
611                   119  English  70.000    1997.000       7.000
683                  2968  English  63.000    1999.000       8.800
792                    91  English  60.000    2003.000       6.700
940                   406  English  60.000    1994.000       7.600
1490                  975  English  32.000    2011.000       6.700
1722                  415  English  30.000    2007.000       7.500
2204                  908  English  25.000    2006.000       7.500
2333                   61  English  10.000    2015.000       5.300
2682                  369  English  15.000    2012.000       6.200
2898                  460  English  13.000    1993.000       8.000

      movie_facebook_likes  profit    genre_1    genre_2
410                      0  27.700     Comedy      Drama
1106                     0   1.820     Action  Adventure
1204                 13000  54.130  Biography      Drama
1408                     0  89.730     Comedy      Drama
```

```
1483                 0 -20.000     Drama   Thriller
1575                 0  56.100 Biography      Drama
1618                 0  33.540    Comedy      Drama
1674               592  -6.790     Drama      Drama
1925                 0  16.600     Drama    Romance
2781             18000  16.960 Biography      Drama
3135               683  10.340    Comedy      Drama
26               26000 458.670     Drama    Romance
50              115000  39.810     Drama    Romance
97              175000 132.570    Action  Adventure
179             190000  48.640 Adventure      Drama
257                  0  -7.390 Biography      Drama
296             199000  62.800     Drama    Western
307              14000 -42.630 Adventure      Drama
308             138000  16.870 Biography     Comedy
326                  0 -22.320     Crime      Drama
361              29000  42.370     Crime      Drama
452              53000  47.970   Mystery   Thriller
641                  0 -30.620    Action      Drama
911              15000 112.440 Biography      Crime
990                  0 -10.220 Adventure      Drama
1114                 0 -12.120     Drama    Romance
1422                 0  21.880    Action  Adventure
1453             16000   2.300 Biography      Crime
1560                 0 -13.360    Action   Thriller
2067              1000 -10.220     Drama      Drama
2757             10000  31.840     Drama    Romance
3476            115000  39.810     Drama    Romance
101              23000 -22.510     Drama    Fantasy
147                  0 -41.770 Adventure  Adventure
254                  0  15.530     Crime   Thriller
255                  0  66.340    Action     Comedy
382                  0 -91.970    Action      Crime
400                  0  98.410     Crime   Thriller
470              82000  17.710    Action      Drama
611                  0 -32.100 Adventure  Biography
683              48000 -25.980     Drama      Drama
792                880 -33.710 Adventure  Animation
940              11000  45.260     Drama    Fantasy
1490             39000 -18.700     Drama    Fantasy
1722                 0 -26.100 Biography      Crime
2204                 0   9.300     Drama      Drama
2333                 0  -9.470     Drama    Romance
2682             20000  -0.060     Crime   Thriller
2898             15000  -0.720    Action      Crime
```

In [229…

```
# Write your code for grouping the combined dataframe here

movies_by_lead_actor = Combined.groupby(['actor_1_name'])

movies_by_lead_actor
```

Out[229…

	num_critic_for_reviews	num_user_for_reviews
actor_1_name		
Brad Pitt	245.000	483629095752844826500841764096304257930765281 52...
Leonardo DiCaprio	330.190	12041682287199465949590673927567270018268705975...
Meryl Streep	181.455	19517525119362000835922425282 56.000

In [230…

```
# Write the code for finding the mean of critic reviews and audience reviews here
```

```
print(movies_by_lead_actor.num_critic_for_reviews.mean())
print(movies_by_lead_actor.num_user_for_reviews.mean())
```

```
actor_1_name
Brad Pitt            245.000
Leonardo DiCaprio    330.190
Meryl Streep         181.455
Name: num_critic_for_reviews, dtype: float64
actor_1_name
Brad Pitt            4836290957528448265008417640963042579307652815 2...
Leonardo DiCaprio    1204168228719946594959067392756727001826870597 5...
Meryl Streep                              19517525119362000835922425282 56.000
Name: num_user_for_reviews, dtype: float64
```

Checkpoint 6: Leonardo has aced both the lists!

In [232…

```
Movies_by_decade=actual_data.copy(deep=True)
Movies_by_decade['decade']=Movies_by_decade['title_year'].apply(lambda x:10*(int(x/10)))
Movies_by_decade

# In this part we are trying to understand how movies have changed in each Decade.
```

Out[232…

	director_name	num_critic_for_reviews	gross	genres	actor_1_name	movie_title
0	James Cameron	723.000	760.510	Action|Adventure|Fantasy|Sci-Fi	CCH Pounder	Avatar
1	Gore Verbinski	302.000	309.400	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End
2	Sam Mendes	602.000	200.070	Action|Adventure|Thriller	Christoph Waltz	Spectre
3	Christopher Nolan	813.000	448.130	Action|Thriller	Tom Hardy	The Dark Knight Rises
5	Andrew Stanton	462.000	73.060	Action|Adventure|Sci-Fi	Daryl Sabara	John Carter
...	...	...	...	...	...	...
5033	Shane Carruth	143.000	0.420	Drama|Sci-Fi|Thriller	Shane Carruth	Primer
5034	Neill Dela Llana	35.000	0.070	Thriller	Ian Gamazon	Cavite
5035	Robert Rodriguez	56.000	2.040	Action|Crime|Drama|Romance|Thriller	Carlos Gallardo	El Mariachi
5037	Edward Burns	14.000	0.000	Comedy|Drama	Kerry Bishé	Newlyweds
5042	Jon Gunn	43.000	0.090	Documentary	John August	My Date with Drew

3856 rows × 17 columns

In [233…

```
# Write your code for creating the data frame df_by_decade here

Movies_by_decade = Movies_by_decade.groupby('decade',as_index=False)['num_voted_users'].su
Movies_by_decade
```

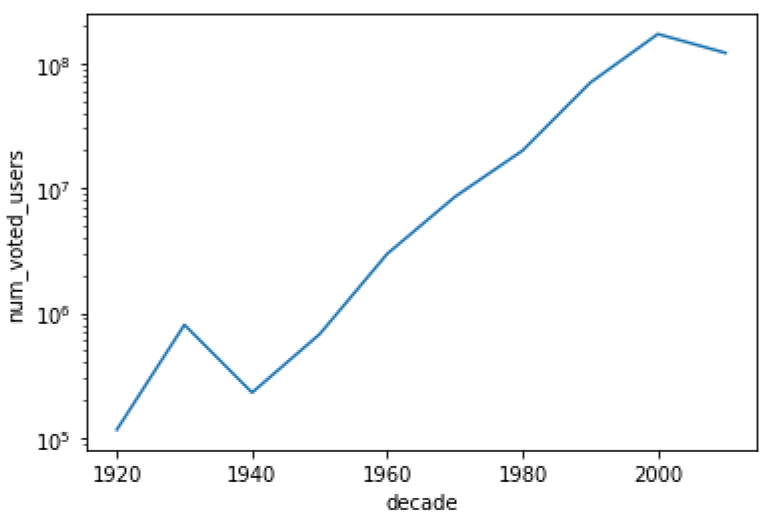
```
# As we can see, the number of Users voted in the IMDB site has been increasing rapidly a
```

Out[233…

	decade	num_voted_users
0	1920	116392
1	1930	804839
2	1940	230838
3	1950	678336
4	1960	2983442
5	1970	8524102
6	1980	19987476
7	1990	69735679
8	2000	170908676
9	2010	120640994

In [242…

```
# Write your code for plotting number of voted users vs decade

sns.lineplot(data= Movies_by_decade, x='decade',y='num_voted_users')
plt.yscale('log')
plt.ylabel("num_voted_users")
plt.xlabel("decade")
plt.show()

# Though the number of Users voted has been since 1900s, but there has been slight dip in
```



In [ ]:

```
```