IMDB MOVIE RATINGS ANALYSIS

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

```
from numpy import corrcoef
import numpy.ma
```

```
# Displaying the data
!head -10 /home/jovian/./titanic_top_1000.csv"
data_head(5)
```

```
Poster_Link,Series_Title,Released_Year,Certificate,Runtime,Genre,IMDB_Rating,Overview,Meta_score,Director,Star1,Star2,Star3,Star4,No_of_Votes,Gross
https://m.media-amazon.com/images/M/MV5BMDFkYTc5MGUtNjcoZi00NTI5LTgxMjctNTRhNjAwNjI1ZjRlXkEyXkFqcGdeQXVyMjAwNjU5NjY@._V1_UY268_CR0,0,182,268_AL_.jpg,The Shawshank Redemption,1994,A,142 min,Drama,9.3,"Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.",80.0,Frank Darabont,Tim Robbins,Morgan Freeman,Bob Gunton,William Sadler,2343110,28,341,903,645,1305037,28,887,541,994,134936924
https://m.media-amazon.com/images/M/MV5BMTMxNTMyODIzNF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg,The Godfather,1972,A,175 min,Crime, Drama,9.2,"An organized crime dynasty aging patriarch passes the torch to the third and reluctant son.",100.0,Francis Ford Coppola,Marlon Brando,Al Pacino,James Caan,Diane Keaton,1620367,11,575,773,8000,534858444,
https://m.media-amazon.com/images/M/MV5BMTQ5NzMyMjgzMF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg,The Dark Knight,2008,UA,152 min,Drama, Action, Crime,9.0,"When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, the Dark Knight must accept one of the greatest psychological and physical tests of his ability to fight injustice.",84.0,Christopher Nolan,Christian Bale,Heath Ledger,Aaron Eckhart,Michael Caine,2303232,9,670,948,12000,534858444,
```

```
data_shape  #Determining the dimension of data
```

```
(1000, 16)
```

```
data_columns  #Brief view of columns headings
```

```
Index(['Poster_Link', 'Series_Title', 'Released_Year', 'Certificate',
       'Runtime', 'Genre', 'IMDB_Rating', 'Overview', 'Meta_score',
       'Director', 'Star1', 'Star2', 'Star3', 'Star4', 'No_of_Votes',
       'Gross'],
      dtype=object)
```

Data Cleaning

```
data_isnull().sum()  #To determine if any null values in the data
```

```
Poster_Link    0
Series_Title   0
Released_Year  0
Certificate    0
Runtime        0
Genre          0
IMDB_Rating    0
Overview       0
Meta_score     0
Director       0
Star1          0
Star2          0
Star3          0
Star4          0
No_of_Votes    0
Gross          0
dtype: int64
```

```
#Filling the null values in null columns
data['Meta_score'].fillna(0,inplace=True)
data['Certificate'].fillna('Unclassified',inplace=True)
data['Gross'].fillna(0,inplace=True)
```

```
data_isnull().sum()  #Rechecking for the null values
```

```
Poster_Link    0
Series_Title   0
Released_Year  0
Certificate    0
Runtime        0
Genre          0
IMDB_Rating    0
Overview       0
Meta_score     0
Director       0
Star1          0
Star2          0
Star3          0
Star4          0
No_of_Votes    0
Gross          0
dtype: int64
```

Use of one of the null values are removed from the columns.Next step we will perform to to standardize each columns and thus cleaning the data.

```
!head data_head(5)  #Rechecking if we have the original values
data_head(5)
```

	Poster_Link	Series_Title	Released_Year	Certificate	Runtime	Genre	IMDB_Rating	Overview	Meta_score	Director	Star1	Star2	Star3	Star4	No_of_Votes	Gross
0	https://m.media-amazon.com/images/M/MV5BMDFkYTc5MGUtNjcoZi00NTI5LTgxMjctNTRhNjAwNjI1ZjRlXkEyXkFqcGdeQXVyMjAwNjU5NjY@._V1_UY268_CR0,0,182,268_AL_.jpg	The Shawshank Redemption	1994	A	142 min	Drama	9.3	Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.	80.0	Frank Darabont	Tim Robbins	Morgan Freeman	Bob Gunton	William Sadler	2343110	28,341,903,645,1305037,28,887,541,994,134936924
1	https://m.media-amazon.com/images/M/MV5BMTMxNTMyODIzNF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Godfather	1972	A	175 min	Crime, Drama	9.2	An organized crime dynasty aging patriarch passes the torch to the third and reluctant son.	100.0	Francis Ford Coppola	Marlon Brando	Al Pacino	James Caan	Diane Keaton	1620367	11,575,773,8000,534858444,0
2	https://m.media-amazon.com/images/M/MV5BMTQ5NzMyMjgzMF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Dark Knight	2008	UA	152 min	Action, Crime, Drama	9.0	When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, the Dark Knight must accept one of the greatest psychological and physical tests of his ability to fight injustice.	84.0	Christopher Nolan	Christian Bale	Heath Ledger	Aaron Eckhart	Michael Caine	2303232	9,670,948,12000,534858444,0

```
data_shape  #Rechecking the dimension of data
```

```
(1000, 16)
```

```
data_columns  #Brief view of columns headings
```

```
Index(['Poster_Link', 'Series_Title', 'Released_Year', 'Certificate',
       'Runtime', 'Genre', 'IMDB_Rating', 'Overview', 'Meta_score',
       'Director', 'Star1', 'Star2', 'Star3', 'Star4', 'No_of_Votes',
       'Gross'],
      dtype=object)
```

Data Cleaning

```
data_isnull().sum()  #To determine if any null values in the data
```

```
Poster_Link    0
Series_Title   0
Released_Year  0
Certificate    0
Runtime        0
Genre          0
IMDB_Rating    0
Overview       0
Meta_score     0
Director       0
Star1          0
Star2          0
Star3          0
Star4          0
No_of_Votes    0
Gross          0
dtype: int64
```

Use of one of the null values are removed from the columns.Next step we will perform to to standardize each columns and thus cleaning the data.

```
!head data_head(5)  #Rechecking if we have the original values
data_head(5)
```

	Poster_Link	Series_Title	Released_Year	Certificate	Runtime	Genre	IMDB_Rating	Overview	Meta_score	Director	Star1	Star2	Star3	Star4	No_of_Votes	Gross
0	https://m.media-amazon.com/images/M/MV5BMDFkYTc5MGUtNjcoZi00NTI5LTgxMjctNTRhNjAwNjI1ZjRlXkEyXkFqcGdeQXVyMjAwNjU5NjY@._V1_UY268_CR0,0,182,268_AL_.jpg	The Shawshank Redemption	1994	A	142 min	Drama	9.3	Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.	80.0	Frank Darabont	Tim Robbins	Morgan Freeman	Bob Gunton	William Sadler	2343110	28,341,903,645,1305037,28,887,541,994,134936924
1	https://m.media-amazon.com/images/M/MV5BMTMxNTMyODIzNF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Godfather	1972	A	175 min	Crime, Drama	9.2	An organized crime dynasty aging patriarch passes the torch to the third and reluctant son.	100.0	Francis Ford Coppola	Marlon Brando	Al Pacino	James Caan	Diane Keaton	1620367	11,575,773,8000,534858444,0
2	https://m.media-amazon.com/images/M/MV5BMTQ5NzMyMjgzMF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Dark Knight	2008	UA	152 min	Action, Crime, Drama	9.0	When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, the Dark Knight must accept one of the greatest psychological and physical tests of his ability to fight injustice.	84.0	Christopher Nolan	Christian Bale	Heath Ledger	Aaron Eckhart	Michael Caine	2303232	9,670,948,12000,534858444,0

The reason to perform the step is to remove string elements from the column. ,thus we can use "Runtime" column during further analysis part

```
data_drop('Runtime',1, axis= 1, inplace= True)  #To drop the unnecessary column
data_head(5)
```

```
Index(['Poster_Link', 'Series_Title', 'Released_Year', 'Certificate',
       'IMDB_Rating', 'Overview', 'Meta_score', 'Director', 'Star1',
       'Star2', 'Star3', 'Star4', 'No_of_Votes', 'Gross'],
      dtype=object)
```

Here we had removed the column "Runtime" it serves as not successfully exponential the integer part from it,, that are displayed in "0" column. Now we can further change the dimensions from minutes to hours as better.

```
data.rename(columns={'0':'Runtime'}, inplace= True)  #Renaming the '0' column
data['Runtime'] = data['Runtime'].astype(float).round(2)  #Converting the minutes to hours
data_head(5)
```

	Poster_Link	Series_Title	Released_Year	Certificate	Runtime	Genre	IMDB_Rating	Overview	Meta_score	Director	Star1	Star2	Star3	Star4	No_of_Votes	Gross
0	https://m.media-amazon.com/images/M/MV5BMDFkYTc5MGUtNjcoZi00NTI5LTgxMjctNTRhNjAwNjI1ZjRlXkEyXkFqcGdeQXVyMjAwNjU5NjY@._V1_UY268_CR0,0,182,268_AL_.jpg	The Shawshank Redemption	1994	A	142 min	Drama	9.3	Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.	80.0	Frank Darabont	Tim Robbins	Morgan Freeman	Bob Gunton	William Sadler	2343110	28,341,903,645,1305037,28,887,541,994,134936924
1	https://m.media-amazon.com/images/M/MV5BMTMxNTMyODIzNF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Godfather	1972	A	175 min	Crime, Drama	9.2	An organized crime dynasty aging patriarch passes the torch to the third and reluctant son.	100.0	Francis Ford Coppola	Marlon Brando	Al Pacino	James Caan	Diane Keaton	1620367	11,575,773,8000,534858444,0
2	https://m.media-amazon.com/images/M/MV5BMTQ5NzMyMjgzMF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Dark Knight	2008	UA	152 min	Action, Crime, Drama	9.0	When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, the Dark Knight must accept one of the greatest psychological and physical tests of his ability to fight injustice.	84.0	Christopher Nolan	Christian Bale	Heath Ledger	Aaron Eckhart	Michael Caine	2303232	9,670,948,12000,534858444,0

```
data.rename(columns={'Runtime':'Runtime(hours)'},inplace=True)
data_columns
```

```
Index(['Poster_Link', 'Series_Title', 'Released_Year', 'Certificate', 'Genre',
       'IMDB_Rating', 'Overview', 'Meta_score', 'Director', 'Star1',
       'Star2', 'Star3', 'Star4', 'No_of_Votes', 'Gross', 'Runtime',
       'Runtime(hours)'],
      dtype=object)
```

Here we had added the new column to the data and removed the old one with comma values. Now we can rename the column.

```
data.rename(columns={'Runtime':'Runtime(hours)'},inplace=True)
data_columns
```

```
Index(['Poster_Link', 'Series_Title', 'Released_Year', 'Certificate', 'Genre',
       'IMDB_Rating', 'Overview', 'Meta_score', 'Director', 'Star1',
       'Star2', 'Star3', 'Star4', 'No_of_Votes', 'Gross', 'Runtime',
       'Runtime(hours)'],
      dtype=object)
```

```
#Rechecking the new column
data['Runtime(hours)'].astype(int).sum()  #Rechecking the summation value to avoid repetations.
```

```
data_head(5)
```

	Poster_Link	Series_Title	Released_Year	Certificate	Genre	IMDB_Rating	Overview	Meta_score	Director	Star1	Star2	Star3	Star4	No_of_Votes	Gross	Runtime	Runtime(hours)
0	https://m.media-amazon.com/images/M/MV5BMDFkYTc5MGUtNjcoZi00NTI5LTgxMjctNTRhNjAwNjI1ZjRlXkEyXkFqcGdeQXVyMjAwNjU5NjY@._V1_UY268_CR0,0,182,268_AL_.jpg	The Shawshank Redemption	1994	A	Drama	9.3	Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.	80.0	Frank Darabont	Tim Robbins	Morgan Freeman	Bob Gunton	William Sadler	2343110	28,341,903,645,1305037,28,887,541,994,134936924	2.37	2.37
1	https://m.media-amazon.com/images/M/MV5BMTMxNTMyODIzNF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Godfather	1972	A	Crime, Drama	9.2	An organized crime dynasty aging patriarch passes the torch to the third and reluctant son.	100.0	Francis Ford Coppola	Marlon Brando	Al Pacino	James Caan	Diane Keaton	1620367	11,575,773,8000,534858444,0	2.92	2.92
2	https://m.media-amazon.com/images/M/MV5BMTQ5NzMyMjgzMF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Dark Knight	2008	UA	Action, Crime, Drama	9.0	When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, the Dark Knight must accept one of the greatest psychological and physical tests of his ability to fight injustice.	84.0	Christopher Nolan	Christian Bale	Heath Ledger	Aaron Eckhart	Michael Caine	2303232	9,670,948,12000,534858444,0	2.52	2.52

Here we had added the new column to the data and removed the old one with comma values. Now we can rename the column.

```
data.rename(columns={'Runtime':'Runtime(hours)'},inplace=True)
data_columns
```

```
Index(['Poster_Link', 'Series_Title', 'Released_Year', 'Certificate', 'Genre',
       'IMDB_Rating', 'Overview', 'Meta_score', 'Director', 'Star1',
       'Star2', 'Star3', 'Star4', 'No_of_Votes', 'Gross', 'Runtime',
       'Runtime(hours)'],
      dtype=object)
```

```
#Rechecking the new column
data.Series_Title.count()  #Rechecking if we have the original values
```

```
1000
```

```
data['avg_rating'][data['IMDB_Rating']>9].sum()  #Calculating the average rating of movies
```

```
9.3
```

```
data_columns  #Any rating values has been inserted here or below.
```

```
Index(['Poster_Link', 'Series_Title', 'Released_Year', 'Certificate', 'Genre',
       'IMDB_Rating', 'Overview', 'Meta_score', 'Director', 'Star1',
       'Star2', 'Star3', 'Star4', 'No_of_Votes', 'Gross', 'Runtime',
       'Runtime(hours)', 'avg_rating'],
      dtype=object)
```

```
#Calculating mean as the average value and round value with minimum to simplify further calculations
data.No_of_Votes[ data.No_of_Votes.notnull()].mean().round(2).astype(float)
data_head(5)
```

```
1272582.0
```

	Poster_Link	Series_Title	Released_Year	Certificate	Genre	IMDB_Rating	Overview	Meta_score	Director	Star1	Star2	Star3	Star4	No_of_Votes	Gross	Runtime	Runtime(hours)	avg_rating
0	https://m.media-amazon.com/images/M/MV5BMDFkYTc5MGUtNjcoZi00NTI5LTgxMjctNTRhNjAwNjI1ZjRlXkEyXkFqcGdeQXVyMjAwNjU5NjY@._V1_UY268_CR0,0,182,268_AL_.jpg	The Shawshank Redemption	1994	A	Drama	9.3	Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.	80.0	Frank Darabont	Tim Robbins	Morgan Freeman	Bob Gunton	William Sadler	2343110	28,341,903,645,1305037,28,887,541,994,134936924	2.37	2.37	9.3
1	https://m.media-amazon.com/images/M/MV5BMTMxNTMyODIzNF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Godfather	1972	A	Crime, Drama	9.2	An organized crime dynasty aging patriarch passes the torch to the third and reluctant son.	100.0	Francis Ford Coppola	Marlon Brando	Al Pacino	James Caan	Diane Keaton	1620367	11,575,773,8000,534858444,0	2.92	2.92	9.2
2	https://m.media-amazon.com/images/M/MV5BMTQ5NzMyMjgzMF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Dark Knight	2008	UA	Action, Crime, Drama	9.0	When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, the Dark Knight must accept one of the greatest psychological and physical tests of his ability to fight injustice.	84.0	Christopher Nolan	Christian Bale	Heath Ledger	Aaron Eckhart	Michael Caine	2303232	9,670,948,12000,534858444,0	2.52	2.52	9.0

Above table shows both IMDB rating and meta score in a scale of 10 as well as Gross, let column connected to billions values

```
#Calculating mean as the average value and round value with minimum to simplify further calculations
data['avg_rating'][data['IMDB_Rating']>9].sum().round(2)
```

```
9.3
```

```
data_columns  #Any rating values has been inserted here or below.
```

```
Index(['Poster_Link', 'Series_Title', 'Released_Year', 'Certificate', 'Genre',
       'IMDB_Rating', 'Overview', 'Meta_score', 'Director', 'Star1',
       'Star2', 'Star3', 'Star4', 'No_of_Votes', 'Gross', 'Runtime',
       'Runtime(hours)', 'avg_rating'],
      dtype=object)
```

```
#Calculating mean as the average value and round value with minimum to simplify further calculations
data.No_of_Votes[ data.No_of_Votes.notnull()].mean().round(2).astype(float)
data_head(5)
```

```
1272582.0
```

	Poster_Link	Series_Title	Released_Year	Certificate	Genre	IMDB_Rating	Overview	Meta_score	Director	Star1	Star2	Star3	Star4	No_of_Votes	Gross	Runtime	Runtime(hours)	avg_rating
0	https://m.media-amazon.com/images/M/MV5BMDFkYTc5MGUtNjcoZi00NTI5LTgxMjctNTRhNjAwNjI1ZjRlXkEyXkFqcGdeQXVyMjAwNjU5NjY@._V1_UY268_CR0,0,182,268_AL_.jpg	The Shawshank Redemption	1994	A	Drama	9.3	Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.	80.0	Frank Darabont	Tim Robbins	Morgan Freeman	Bob Gunton	William Sadler	2343110	28,341,903,645,1305037,28,887,541,994,134936924	2.37	2.37	9.3
1	https://m.media-amazon.com/images/M/MV5BMTMxNTMyODIzNF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Godfather	1972	A	Crime, Drama	9.2	An organized crime dynasty aging patriarch passes the torch to the third and reluctant son.	100.0	Francis Ford Coppola	Marlon Brando	Al Pacino	James Caan	Diane Keaton	1620367	11,575,773,8000,534858444,0	2.92	2.92	9.2
2	https://m.media-amazon.com/images/M/MV5BMTQ5NzMyMjgzMF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Dark Knight	2008	UA	Action, Crime, Drama	9.0	When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, the Dark Knight must accept one of the greatest psychological and physical tests of his ability to fight injustice.	84.0	Christopher Nolan	Christian Bale	Heath Ledger	Aaron Eckhart	Michael Caine	2303232	9,670,948,12000,534858444,0	2.52	2.52	9.0

Now the data analysis part will begin from here.

1. Top 3 movies on the basis of average ratings. Sort them on the basis of avg_rating and IMDB_Rating

```
data.sort_values(by=['avg_rating','IMDB_Rating'], ascending= False,inplace= False).head(3)
```

	Poster_Link	Series_Title	Released_Year	Certificate	Genre	IMDB_Rating	Overview	Meta_score	Director	Star1	Star2	Star3	Star4	No_of_Votes	Gross	Runtime	Runtime(hours)	avg_rating
0	https://m.media-amazon.com/images/M/MV5BMDFkYTc5MGUtNjcoZi00NTI5LTgxMjctNTRhNjAwNjI1ZjRlXkEyXkFqcGdeQXVyMjAwNjU5NjY@._V1_UY268_CR0,0,182,268_AL_.jpg	The Shawshank Redemption	1994	A	Drama	9.3	Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.	80.0	Frank Darabont	Tim Robbins	Morgan Freeman	Bob Gunton	William Sadler	2343110	28,341,903,645,1305037,28,887,541,994,134936924	2.37	2.37	9.3
1	https://m.media-amazon.com/images/M/MV5BMTMxNTMyODIzNF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Godfather	1972	A	Crime, Drama	9.2	An organized crime dynasty aging patriarch passes the torch to the third and reluctant son.	100.0	Francis Ford Coppola	Marlon Brando	Al Pacino	James Caan	Diane Keaton	1620367	11,575,773,8000,534858444,0	2.92	2.92	9.2
2	https://m.media-amazon.com/images/M/MV5BMTQ5NzMyMjgzMF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Dark Knight	2008	UA	Action, Crime, Drama	9.0	When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, the Dark Knight must accept one of the greatest psychological and physical tests of his ability to fight injustice.	84.0	Christopher Nolan	Christian Bale	Heath Ledger	Aaron Eckhart	Michael Caine	2303232	9,670,948,12000,534858444,0	2.52	2.52	9.0

1. Top 3 movies on the basis of number of votes.

```
data.sort_values(by=['No_of_Votes'],ascending= False,inplace=False).head(3)
```

	Poster_Link	Series_Title	Released_Year	Certificate	Genre	IMDB_Rating	Overview	Meta_score	Director	Star1	Star2	Star3	Star4	No_of_Votes	Gross	Runtime	Runtime(hours)	avg_rating
0	https://m.media-amazon.com/images/M/MV5BMDFkYTc5MGUtNjcoZi00NTI5LTgxMjctNTRhNjAwNjI1ZjRlXkEyXkFqcGdeQXVyMjAwNjU5NjY@._V1_UY268_CR0,0,182,268_AL_.jpg	The Shawshank Redemption	1994	A	Drama	9.3	Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.	80.0	Frank Darabont	Tim Robbins	Morgan Freeman	Bob Gunton	William Sadler	2343110	28,341,903,645,1305037,28,887,541,994,134936924	2.37	2.37	9.3
1	https://m.media-amazon.com/images/M/MV5BMTMxNTMyODIzNF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Godfather	1972	A	Crime, Drama	9.2	An organized crime dynasty aging patriarch passes the torch to the third and reluctant son.	100.0	Francis Ford Coppola	Marlon Brando	Al Pacino	James Caan	Diane Keaton	1620367	11,575,773,8000,534858444,0	2.92	2.92	9.2
2	https://m.media-amazon.com/images/M/MV5BMTQ5NzMyMjgzMF5BMl5BanBnXkFtZTcwMjcyNjcyNg@@._V1_UY268_CR0,0,182,268_AL_.jpg	The Dark Knight	2008	UA	Action, Crime, Drama	9.0	When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, the Dark Knight must accept one of the greatest psychological and physical tests of his ability to fight injustice.	84.0	Christopher Nolan	Christian Bale	Heath Ledger	Aaron Eckhart	Michael Caine	2303232	9,670,948,12000,534858444,0	2.52	2.52	9.0

Number of movies in the particular category

```
data.Certificate.value_counts()
```

```
Certificate
A             234
UA            197
U             176
R             167
PG-13          43
PG             39
G              14
Passed         13
Unclassified   11
TV-PG           6
TV-14           4
TV-MA           4
Approved        2
Name: count, dtype: int64
```

Category wise average gross earnings

```
data.groupby('Certificate').Gross.mean().sort_values(ascending= False).round(2)  #All values are in Billions
```

```
Certificate
TV-14          0.00
TV-MA          0.00
G              0.00
U              0.00
PG             0.00
PG-13          0.00
R              0.00
A              0.00
Passed         0.00
UA             0.00
TV-PG          0.00
Approved       0.00
Unclassified   0.00
Name: Gross, dtype: float64
```

Category wise number of votes.

```
data.groupby('Certificate').No_of_Votes.sum().sort_values(ascending= False)  #All values are in Billions
```

```
Certificate
A             0.00
UA            0.00
U             0.00
PG-13         0.00
PG            0.00
R             0.00
G             0.00
Passed        0.00
TV-PG         0.00
TV-14         0.00
TV-MA         0.00
Approved      0.00
Unclassified  0.00
Name: No_of_Votes, dtype: float64
```

Number of movies by each director.....show only top of top 20

```
data.Director.value_counts().sort_values(ascending= False).head(20)
```

```
Director
Alfred Hitchcock    14
Steven Spielberg    13
Hayao Miyazaki       9
Martin Scorsese      9
Stanley Kubrick      9
Akira Kurosawa       8
Billy Wilder         8
Quentin Tarantino    8
David Fincher        7
Christopher Nolan    7
Ridley Scott         7
Tim Burton           7
Ang Lee              6
Bob Rafelson         6
Clint Eastwood       6
Robert Bresson       6
Sergio Leone         6
Richard Linklater    5
Terence Malick       5
Wes Anderson         5
Woody Allen          5
Yasujirô Ozu         5
Name: count, dtype: int64
```

Average Gross earnings per movies of directors...only show top directors

```
data.groupby('Director').Gross.mean().round(2)
```

```
Director
Christopher Nolan    0.00
Clint Eastwood       0.00
Martin Scorsese      0.00
Steven Spielberg     0.00
Akira Kurosawa       0.00
Robert Bresson       0.00
Quentin Tarantino    0.00
Richard Linklater    0.00
Terence Malick       0.00
Wes Anderson         0.00
Woody Allen          0.00
Yasujirô Ozu         0.00
Sergio Leone         0.00
Asghar Farhadi       0.00
Stanley Kubrick      0.00
David Fincher        0.00
Ang Lee              0.00
Billy Wilder         0.00
Alfred Hitchcock     0.00
Hayao Miyazaki       0.00
Bob Rafelson         0.00
Name: Gross, dtype: float64
```

Total number of votes each director movies received

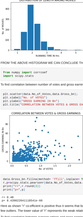
```
data.groupby('Director').No_of_Votes.sum().round(2).sort_values(ascending= False)
```

```
Director
Christopher Nolan    0.00
Quentin Tarantino    0.00
Martin Scorsese      0.00
Steven Spielberg     0.00
David Fincher        0.00
Akira Kurosawa       0.00
Robert Bresson       0.00
Clint Eastwood       0.00
Billy Wilder         0.00
Stanley Kubrick      0.00
Ridley Scott         0.00
Hayao Miyazaki       0.00
Tim Burton           0.00
Ang Lee              0.00
Alfred Hitchcock     0.00
Sergio Leone         0.00
Wes Anderson         0.00
Woody Allen          0.00
Yasujirô Ozu         0.00
Bob Rafelson         0.00
Name: No_of_Votes, dtype: float64
```

Data Visualization

Distribution of IMDB RATINGS among movies

```
plt.hist(data.IMDB_Rating, range=(0,9))
plt.xlabel("IMDB RATINGS")
plt.ylabel("No. OF MOVIES")
plt.title("DISTRIBUTION OF RATINGS AMONG MOVIES");
```



THE ABOVE HISTOGRAM SHOWS THAT THIS IS A RIGHT SKEWED PLOT WHICH MEANS THAT MAXIMUM NUMBER OF MOVIES HAVE RATINGS WHICH ARE LESS THAN THE OVERALL MEAN RATINGS.

Distribution of Average ratings among movies

```
plt.hist(data.avg_rating, range=(0,9))
plt.xlabel("AVERAGE RATINGS")
plt.ylabel("No. OF MOVIES")
plt.title("DISTRIBUTION OF AVERAGE RATINGS AMONG MOVIES");
```

THE ABOVE HISTOGRAM IS LEFT SKEWED WHICH SHOWS THAT MAXIMUM NUMBER OF MOVIES RATINGS ARE BETWEEN 7.4 AND 9.9. THUS ACCORDING TO THIS RATING MAXIMUM NUMBER OF MOVIES ARE HAVING THEIR FINAL RATING HIGHER THAN THEIR OVERALL MEAN RATINGS.

Distribution of Running length among movies

```
plt.hist(data.Runtime(hrs), range=(0,9))
plt.xlabel("RUNNING TIME IN HRS")
plt.ylabel("No. OF MOVIES")
plt.title("DISTRIBUTION OF LENGTH AMONG MOVIES");
```

FROM THE ABOVE HISTOGRAM WE CAN CONCLUDE THAT MAXIMUM MOVIES HAVE THEIR TOTAL RUNNING TIME BETWEEN 1.5 AND 2.5 HOURS. AS A RESULT OF WHICH THE GRAPH IS RIGHT SKEWED.

```
import numpy
import matplotlib
%matplotlib inline
```

To find correlation between number of votes and gross earnings

```
plt.scatter(data.No_of_Votes,data.Gross,s=5)
plt.xlabel("No. OF VOTES")
plt.ylabel("GROSS EARNINGS")
plt.title("CORRELATION BETWEEN VOTES & GROSS EARNINGS");
```

```
data.Gross.to_frame(inplace= True).corr(inplace= True)
data.No_of_Votes.to_frame(inplace= True).corr(inplace= True)
print("0.4")
print("0.4")
```

```
0.4
0.4
```

```
# 0.4 indicates positive
```

0.4 indicates positive that it seems that both gross earnings and number of votes are having positive correlation. As a result, higher the number of votes higher will be the gross earnings. But still there are few outliers. The lower value of 'r' represents the weak relations between both of them.

To find correlation between average rate and gross earnings

```
plt.scatter(data.avg_rating,data.Gross,s=5)
plt.xlabel("AVERAGE RATINGS")
plt.ylabel("GROSS EARNINGS")
plt.title("CORRELATION BETWEEN AVG RATINGS & GROSS EARNINGS");
```

```
# or using stats package(scipy.stats.stats)
data.avg_rating.to_frame(inplace= True).corr(inplace= True)
print("0.1")
print("0.1")
```

```
0.1
0.1
```

```
# 0.1 indicates positive
```

0.1 indicates positive that it seems that both gross earnings and average rating are having positive correlation. As a result, higher the number of votes higher will be the gross earnings. But still there are few outliers. The lower value of 'r' represents the weak relations between both of them.

To find correlation between number of votes and the movies.

```
plt.scatter(data.No_of_Votes,data.Series_Title,s=5)
plt.xlabel("NO. OF VOTES")
plt.ylabel("TITLE OF MOVIES")
plt.title("CORRELATION BETWEEN VOTES & NO. OF MOVIES");
```

```
# or using stats package(scipy.stats.stats)
data.No_of_Votes.to_frame(inplace= True).corr(inplace= True)
print("0.1")
print("0.1")
```

```
0.1
0.1
```

```
# 0.1 indicates positive
```

0.1 indicates positive that it seems that both gross earnings and number of votes are having positive correlation. As a result, higher the number of votes higher will be the gross earnings. But still there are few outliers. The lower value of 'r' represents the weak relations between both of them.