

```
In [0]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.figure import Figure
fig=plt.figure(figsize=(12,8))

In [2]: df=pd.read_csv('movies.csv')

In [3]: df.head()

Out[3]:
```

	name	rating	genre	year	released	score	votes	director	writer	star	country	budget	gross	company	runtime
0	The Shawshank Redemption	R	Drama	1994	July 14, 1994 (United States)	8.8	970000.0	Tim Robbins	Stephen King	Morgan Freeman	United States	23700000.0	67896100.0	Warner Bros.	143.0
1	Pulp Fiction	R	Action	1994	April 26, 1994 (United States)	8.9	650000.0	Quentin Tarantino	Howard Shore	Bruce Willis	United States	10000000.0	213998120.0	Columbia Pictures	154.0
2	Star Wars Episode V - The Empire Strikes Back	PG	Action	1980	June 20, 1980 (United States)	8.7	1200000.0	Ivan Kershew	Leigh Brackett	Mark Hamill	United States	18000000.0	538375067.0	Lucasfilm	124.0
3	Alien	PG-13	Horror	1979	May 25, 1979 (United States)	8.1	221000.0	James Cameron	Lawrence Kasdan	Sissy Spacek	United States	11000000.0	103688319.0	Paramount Pictures	97.0
4	Caddyshack	R	Comedy	1980	July 25, 1980 (United States)	7.3	108000.0	Harold Ramis	Brian Doyle-Murray	Chevy Chase	United States	6000000.0	39846344.0	Orion Pictures	98.0

```
In [4]: df.dropna(inplace=True)

In [5]: df.dtypes

Out[5]:
```

	name	rating	genre	year	released	score	votes	director	writer	star	country	budget	gross	company	runtime
0	The Shawshank Redemption	R	Drama	1994	July 14, 1994 (United States)	8.8	970000.0	Tim Robbins	Stephen King	Morgan Freeman	United States	23700000.0	67896100.0	Warner Bros.	143.0
1	Pulp Fiction	R	Action	1994	April 26, 1994 (United States)	8.9	650000.0	Quentin Tarantino	Howard Shore	Bruce Willis	United States	10000000.0	213998120.0	Columbia Pictures	154.0
2	Star Wars Episode V - The Empire Strikes Back	PG	Action	1980	June 20, 1980 (United States)	8.7	1200000.0	Ivan Kershew	Leigh Brackett	Mark Hamill	United States	18000000.0	538375067.0	Lucasfilm	124.0
3	Alien	PG-13	Horror	1979	May 25, 1979 (United States)	8.1	221000.0	James Cameron	Lawrence Kasdan	Sissy Spacek	United States	11000000.0	103688319.0	Paramount Pictures	97.0
4	Caddyshack	R	Comedy	1980	July 25, 1980 (United States)	7.3	108000.0	Harold Ramis	Brian Doyle-Murray	Chevy Chase	United States	6000000.0	39846344.0	Orion Pictures	98.0

```
In [8]: <sort>/>df.sort_values(by='gross',ascending=False,inplace=True)

In [9]: df.head()

Out[9]:
```

	name	rating	genre	year	released	score	votes	director	writer	star	country	budget	gross	company	runtime
5445	Avatar	PG-13	Action	2009	December 18, 2009 (United States)	7.8	1100000.0	James Cameron	James Cameron	Sam Worthington	United States	237000000.0	2847246203.0	Twentieth Century Fox	162.0
7445	Avengers: Endgame	PG-13	Action	2019	April 26, 2019 (United States)	8.4	903000.0	Anthony Russo	Christopher Markus	Robert Downey Jr.	United States	356000000.0	2797501328.0	Marvel Studios	181.0
3045	Titanic	Drama	1997	December 19, 1997 (United States)	7.8	1100000.0	James Cameron	James Cameron	Leonardo DiCaprio	Kate Winslet	United States	200000000.0	2201647264.0	Twentieth Century Fox	194.0
6663	Star Wars Episode VI - The Force Awakens	PG-13	Action	2015	December 18, 2015 (United States)	7.8	876000.0	J.J. Abrams	Lawrence Kasdan	Daisy Ridley	United States	244000000.0	2069621700.0	Lucasfilm	138.0
7244	Avengers: Infinity War	PG-13	Action	2018	April 27, 2018 (United States)	8.4	897000.0	Anthony Russo	Christopher Markus	Robert Downey Jr.	United States	321000000.0	2048359754.0	Marvel Studios	149.0

```
In [18]: pd.set_option("display.max_rows",None)

In [13]: plt.scatter(df['budget'],df['gross'])
plt.title('Budget vs Gross Earnings')
plt.xlabel('Budget')
plt.ylabel('Gross')

Out[11]: Text(0, 0.5, 'Gross')

In [12]: df.head()

Out[12]:
```

	name	rating	genre	year	released	score	votes	director	writer	star	country	budget	gross	company	runtime
5445	Avatar	PG-13	Action	2009	December 18, 2009 (United States)	7.8	1100000.0	James Cameron	James Cameron	Sam Worthington	United States	237000000.0	2847246203.0	Twentieth Century Fox	162.0
7445	Avengers: Endgame	PG-13	Action	2019	April 26, 2019 (United States)	8.4	903000.0	Anthony Russo	Christopher Markus	Robert Downey Jr.	United States	356000000.0	2797501328.0	Marvel Studios	181.0
3045	Titanic	Drama	1997	December 19, 1997 (United States)	7.8	1100000.0	James Cameron	James Cameron	Leonardo DiCaprio	Kate Winslet	United States	200000000.0	2201647264.0	Twentieth Century Fox	194.0
6663	Star Wars Episode VI - The Force Awakens	PG-13	Action	2015	December 18, 2015 (United States)	7.8	876000.0	J.J. Abrams	Lawrence Kasdan	Daisy Ridley	United States	244000000.0	2069621700.0	Lucasfilm	138.0
7244	Avengers: Infinity War	PG-13	Action	2018	April 27, 2018 (United States)	8.4	897000.0	Anthony Russo	Christopher Markus	Robert Downey Jr.	United States	321000000.0	2048359754.0	Marvel Studios	149.0

```
In [13]: sns.scatterplot(x=df['budget'],y=df['gross'])

In [14]: correlation_matrix=df.corr('pearson')
correlation_matrix

Out[14]:
```

	year	score	votes	budget	gross	runtime
year	1.000000	0.056386	0.202001	0.327722		