In [485]:
```
import pandas as pd
from IPython.display import HTML
```

In [486]:
```
# Data Import and first Inspection
```

In [487]:
```
Data=pd.read_csv('movies_complete.csv')
```

Some additional information on Features/Columns:

- **id:** The ID of the movie (clear/unique identifier).
- **title:** The Official Title of the movie.
- **tagline:** The tagline of the movie.
- **release_date:** Theatrical Release Date of the movie.
- **genres:** Genres associated with the movie.
- **belongs_to_collection:** Gives information on the movie series/franchise the particular film belongs to.
- **original_language:** The language in which the movie was originally shot in.
- **budget_musd:** The budget of the movie in million dollars.
- **revenue_musd:** The total revenue of the movie in million dollars.
- **production_companies:** Production companies involved with the making of the movie.
- **production_countries:** Countries where the movie was shot/produced in.
- **vote_count:** The number of votes by users, as counted by TMDB.
- **vote_average:** The average rating of the movie.
- **popularity:** The Popularity Score assigned by TMDB.
- **runtime:** The runtime of the movie in minutes.
- **overview:** A brief blurb of the movie.
- **spoken_languages:** Spoken languages in the film.
- **poster_path:** The URL of the poster image.
- **cast:** (Main) Actors appearing in the movie.
- **cast_size:** number of Actors appearing in the movie.
- **director:** Director of the movie.
- **crew_size:** Size of the film crew (incl. director, excl. actors).

In [488]:
```
Data.head(3)
```

Out[488]:

	id	title	tagline	release_date	genres	belongs_to_collection	origina
0	862	Toy Story	NaN	1995-10-30	Animation|Comedy|Family	Toy Story Collection	
1	8844	Jumanji	Roll the dice and unleash the excitement!	1995-12-15	Adventure|Fantasy|Family	NaN	
2	15602	Grumpier Old Men	Still Yelling. Still Fighting. Still Ready for...	1995-12-22	Romance|Comedy	Grumpy Old Men Collection	

3 rows × 22 columns

In [489]:
```
#Data=Data.set_index('title')
```

In [490]:
```
Data.isnull().sum()
```

Out[490]:
```
id                          0
title                       0
tagline                 24407
release_date               34
genres                   2105
belongs_to_collection   40228
original_language          10
budget_musd             35837
revenue_musd            37306
production_companies    11335
production_countries     5856
vote_count                  0
vote_average             2614
popularity                  0
runtime                  1512
overview                  951
spoken_languages         3597
poster_path               224
cast                     2189
cast_size                   0
crew_size                   0
director                  731
dtype: int64
```

In [491]:
```
Data.shape
```

Out[491]:
```
(44691, 22)
```

In [492]:
```
Data['profit']=Data['revenue_musd']-Data['budget_musd']
Data['ROI']=(Data['revenue_musd']/Data['budget_musd'])*100
```

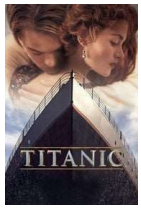
In [ ]:
```

```

In [493]:
```
#_function

def filter_data(Basis,Number_Of_Record,ascending,min_bud=0,minRating=0):
    B=Basis
    N=Number_Of_Record
    A=ascending
    subData=Data[(Data['budget_musd']>=min_bud)&(Data['vote_average']>=minRati
    Record=subData.sort_values(by=B,ascending=A)[0:N][['title','poster_path',E
    return HTML(Record.to_html(escape=False))
```

In [494]:

```
#__Movies Top 5 - Highest Revenue__
Top_5_movie_highest_Revenue=filter_data('revenue_musd',5,False)
Top_5_movie_highest_Revenue

```

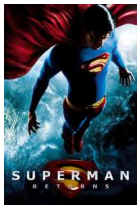
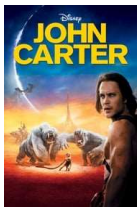
Out[494]:

	poster_path	revenue_musd
title		
Avatar		2787.965087
Star Wars: The Force Awakens		2068.223624
Titanic		1845.034188
The Avengers		1519.557910
Jurassic World		1513.528810

In [495]:

```
#__Movies Top 5 - Highest Budget__
Top_5_movie_highest_Budget=filter_data('budget_musd',5,False)
Top_5_movie_highest_Budget
```

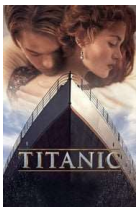
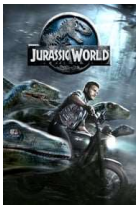
Out[495]:

title	poster_path	budget_musd
Pirates of the Caribbean: On Stranger Tides		380.0
Pirates of the Caribbean: At World's End		300.0
Avengers: Age of Ultron		280.0
Superman Returns		270.0
John Carter		260.0

In [496]:

```
#__Movies Top 5 - Highest Profit__
Top_5_movie_highest_Profit=filter_data('profit',5,False)
Top_5_movie_highest_Profit
```

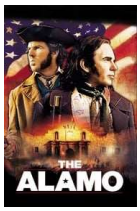
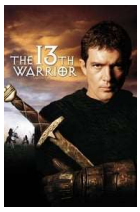
Out[496]:

	poster_path	profit
title		
Avatar		2550.965087
Star Wars: The Force Awakens		1823.223624
Titanic		1645.034188
Jurassic World		1363.528810
Furious 7		1316.249360

In [497]:

```
#__Movies Top 5 - Lowest Profit__
Top_5_movie_Lowest_Profit=filter_data('profit',5,True)
Top_5_movie_Lowest_Profit
```

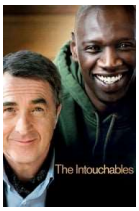
Out[497]:

title	poster_path	profit
The Lone Ranger		-165.710090
The Alamo		-119.180039
Mars Needs Moms		-111.007242
Valerian and the City of a Thousand Planets		-107.447384
The 13th Warrior		-98.301101

In [498]:

```
#__Movies Top 5 - Highest ROI__
#Highest Return on Investment (=Revenue / Budget) (only movies with Budget >
Top_5_movie_Highest_ROI=filter_data('ROI',5,False,10)
Top_5_movie_Highest_ROI
```

Out[498]:

title	poster_path	ROI
E.T. the Extra-Terrestrial		7552.050724
Star Wars		7049.072791
Pretty Woman		3307.142857
The Intouchables		3280.622085
The Empire Strikes Back		2991.111111

In [499]:

```
#Lowest Return on Investment (=Revenue / Budget) (only movies with Budget >=
#__Movies Top 5 - Lowest ROI__
Top_5_movie_Lowest_ROI=filter_data('ROI',5,True,10)
Top_5_movie_Lowest_ROI
```

Out[499]:

title	poster_path	ROI
Chasing Liberty		0.000052
The Cookout		0.000075
Deadfall		0.000180
In the Cut		0.000192
The Samaritan		0.021008

In [500]:

```
#__Movies Top 5 - Most Votes__
#Lowest Rating (only movies with 10 or more Ratings)
Top_5_movie_Most_Votes=filter_data('vote_count',5,False)
Top_5_movie_Most_Votes
```

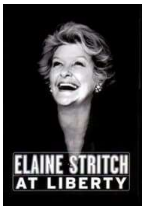
Out[500]:

title	poster_path	vote_count
Inception		14075.0
The Dark Knight		12269.0
Avatar		12114.0
The Avengers		12000.0
Deadpool		11444.0

In [501]:

```
#__Movies Top 5 - Highest Rating__
#Highest Rating (only movies with 10 or more Ratings)
Top_5_movie_Highest_Rating=filter_data('vote_average',5,False,minRating=0)
Top_5_movie_Highest_Rating

```

Out[501]:

title	poster_path	vote_average
Time Pass		10.0
Shuttlecock Boys		10.0
Forever		10.0
Souls of Zen: Ancestors and Agency in Contemporary Japanese Temple Buddhism		10.0
Elaine Stritch: At Liberty		10.0

In [502]:

```
#Lowest Rating (only movies with 10 or more Ratings)
Top_5_movie_Lowest_Rating=filter_data('vote_average',5,True,minRating=0)
Top_5_movie_Lowest_Rating

```

Out[502]:

title	poster_path	vote_average
Extinction: Nature Has Evolved		0.0
Roukli		0.0
Joe and Max		0.0
Call Me by Your Name		0.0
Unrated II: Scary as Hell		0.5

In [503]:
```
#__Movies Top 5 - Most Popular__
Top_5_movie_Most_Popular=filter_data('popularity',5,False)
Top_5_movie_Most_Popular
```

Out[503]:

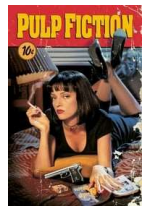
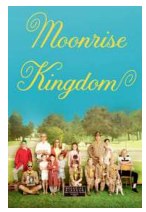
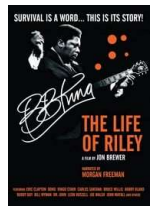
title	poster_path	popularity
Minions		547.488298
Wonder Woman		294.337037
Beauty and the Beast		287.253654
Baby Driver		228.032744
Big Hero 6		213.849907

Find your next Movie

In [504]:
```
#__Search 1: Science Fiction Action Movie with Bruce Willis (sorted from hig
```

In [505]:
```
Gn=Data[Data['genres'].str.contains("Action") & Data['genres'].str.contains(
Ac=Data['cast'].str.contains('Bruce Willis')
sub=Data.loc[(Ac & str(Gn)), ["title", "vote_average",'poster_path']].sort_v
HTML(sub.to_html(escape=False))
```

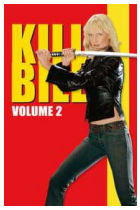
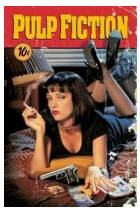
Out[505]:

	title	vote_average	poster_path
291	Pulp Fiction	8.3	
2620	The Sixth Sense	7.7	
18846	Moonrise Kingdom	7.6	
35664	B.B. King: The Life of Riley	7.5	
998	Die Hard	7.5	

In [506]:
```
#__Search 2: Movies with Uma Thurman and directed by Quentin Tarantino (Low
```

In [507]:
```
sub=Data[(Data['cast'].str.contains('Uma Thurman')) & (Data['director']=='Qu
sub=sub[['title','release_date','runtime','poster_path']]
HTML(sub.to_html(escape=False))
```

Out[507]:

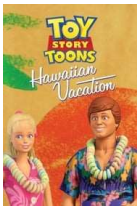
	title	release_date	runtime	poster_path
6667	Kill Bill: Vol. 1	2003-10-10	111.0	
7208	Kill Bill: Vol. 2	2004-04-16	136.0	
291	Pulp Fiction	1994-09-10	154.0	

In [508]:
```
#__Search 3: Most Successful Pixar Studio Movies between 2010 and 2015 (high
```

In [509]:
```
['production_companies']==('Pixar Animation Studios')].sort_values(by='revenue_mu
```

In [510]:
```
sub=sub[sub['release_date'].between("2010-01-01", "2015-12-31")][['title','p
HTML(sub.to_html(escape=False))
```

Out[510]:

	title	production_companies	poster_path
21694	The Blue Umbrella	Pixar Animation Studios	
22489	La luna	Pixar Animation Studios	
24252	Hawaiian Vacation	Pixar Animation Studios	
24254	Small Fry	Pixar Animation Studios	
25516	The Legend of Mor'du	Pixar Animation Studios	

In [511]:
```
#__Search 4: Action or
#Thriller Movie with original language English and minimum Rating of 7.5 (mo
```

In [512]:

```
sub=Data[Data['genres'].str.contains('Action') & Data['genres'].str.contains
sub=sub[(sub['spoken_languages']==('English'))& (sub['vote_average']>=7.5)][
HTML(sub.to_html(escape=False))
```

Out[512]:

title	genres	poster_path	vote_average	release_date
The Dark Knight Rises	Action|Crime|Drama|Thriller		7.6	2012-07-16
Ghost Recon: Alpha	Action|Science Fiction|Thriller|War		7.5	2012-05-03
Oxy-Morons	Action|Thriller		8.0	2011-10-02
Inception	Action|Thriller|Science Fiction|Mystery|Adventure		8.1	2010-07-14
Prison Break: The Final Break	Action|Drama|Thriller		7.5	2009-05-26

```
Data.head(2)
```

In [ ]:

```

```

In [513]:
```
#4. __Analyze__ the Dataset and __find out whether Franchises
#(Movies that belong to a # collection)are more successful
#than stand-alone movies__ in terms of:
```

- mean revenue
- median Return on Investment
- mean budget raised
- mean popularity
- mean rating

In [514]:
```
Data['Franchise']=Data['belongs_to_collection'].notna()
```

In [515]:
```
#__Franchise vs. Stand-alone: Average Revenue__
```

In [516]:
```
Data.groupby(Data['Franchise'])['revenue_musd'].mean().sort_values(ascending
```

Out[516]:
```
Franchise
True     165.708193
False     44.742814
Name: revenue_musd, dtype: float64
```

In [517]:
```
#__Franchise vs. Stand-alone: Return on Investment / Profitability__
```

In [518]:
```
Data.groupby(Data['Franchise'])['ROI'].median().sort_values(ascending=False)
```

Out[518]:
```
Franchise
True     370.919508
False    161.969933
Name: ROI, dtype: float64
```

In [519]:
```
#__Franchise vs. Stand-alone: Average Budget__
Data.groupby(Data['Franchise'])['budget_musd'].mean().sort_values(ascending=
```

Out[519]:
```
Franchise
True     38.319847
False    18.047741
Name: budget_musd, dtype: float64
```

In [520]:
```
#__Franchise vs. Stand-alone: Average Popularity__
Data.groupby(Data['Franchise'])['popularity'].mean().sort_values(ascending=F
```

Out[520]:
```
Franchise
True     6.245051
False    2.592726
Name: popularity, dtype: float64
```

In [521]:
```
#__Franchise vs. Stand-alone: Average Rating__
Data.groupby(Data['Franchise'])['vote_average'].mean().sort_values(ascending
```

Out[521]:
```
Franchise
False    6.008787
True     5.956806
Name: vote_average, dtype: float64
```

Most Successful Franchises

5. **Find** the **most successful Franchises** in terms of

- **total number of movies**
- **total & mean budget**
- **total & mean revenue**
- **mean rating**

In [522]:
```
Fr=Data.groupby(Data['belongs_to_collection']).agg({'title':"count",'budget_
```

In [523]:
```
#total number of movies
Fr.sort_values(by=('title', 'count'),ascending=False).head(1)
```

Out[523]:

	title	budget_musd		revenue_musd		vote_average
	count	sum	mean	sum	mean	mean
belongs_to_collection						
The Bowery Boys	29	0.0	NaN	0.0	NaN	6.675

In [524]:
```
#total & mean budget
Fr.sort_values(by=( 'budget_musd',  'mean'),ascending=False).head(1)
```

Out[524]:

	title	budget_musd		revenue_musd		vote_average
	count	sum	mean	sum	mean	mean
belongs_to_collection						
Tangled Collection	2	260.0	260.0	591.794936	591.794936	7.25

In [525]:
```
#mean rating
Fr.sort_values(by=('vote_average',  'mean'),ascending=False).head(1)
```

Out[525]:

	title	budget_musd		revenue_musd		vote_average
	count	sum	mean	sum	mean	mean
belongs_to_collection						
Argo Collection	1	0.0	NaN	0.0	NaN	9.3

Most Successful Directors

6. **Find** the **most successful Directors** in terms of

- **total number of movies**
- **total revenue**
- **mean rating**

In [526]:
```
Dire=Data.groupby(Data['director']).agg({'title':'count','revenue_musd':'sum
```

In [527]:
```
#- __total number of movies__
Dire.sort_values(by='title',ascending=False).head(1)
```

Out[527]:

	title	revenue_musd	vote_average
director			
John Ford	66	85.170757	6.381818

In [528]:
```
#- __total revenue__
Dire.sort_values(by='revenue_musd',ascending=False).head(1)
```

Out[528]:

	title	revenue_musd	vote_average
director			
Steven Spielberg	33	9256.621422	6.893939

In [529]:
```
#- __mean rating__
Dire.sort_values(by='vote_average',ascending=False).head(1)

```

Out[529]:

	title	revenue_musd	vote_average
director			
Antonis Sotiropoulos	1	0.0	10.0

END

In [ ]:

```
1
```