In [ ]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
import plotly.express as px
```

In [ ]:
```
df= pd.read_csv("C:\\Users\\APURVA BOBADE\\Downloads\\movies.csv",encoding= 'unicode_escape' )
df
```

Out[ ]:

	id	imdb_id	popularity	budget	revenue	original_title	cast	hom
0	135397	tt0369610	32.985763	150000000	1513528810	Jurassic World	Chris Pratt|Bryce Dallas Howard|Irrfan Khan|Vi...	http://www.jurassicworl
1	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...	http://www.madmaxmovi
2	262500	tt2908446	13.112507	110000000	295238201	Insurgent	Shailene Woodley|Theo James|Kate Winslet|Ansel...	http://www.thedivergentseries.movie/#ins
3	140607	tt2488496	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford|Mark Hamill|Carrie Fisher|Adam D...	http://www.starwars.com/films/sta ep
4	168259	tt2820852	9.335014	190000000	1506249360	Furious 7	Vin Diesel|Paul Walker|Jason Statham|Michelle ...	http://www.furious
...	...	...	...	...	...	...	...	
10861	21	tt0060371	0.080598	0	0	The Endless Summer	Michael Hynson|Robert August|Lord 'Tally Ho' B...	
10862	20379	tt0060472	0.065543	0	0	Grand Prix	James Garner|Eva Marie Saint|Yves Montand|Tosh...	
10863	39768	tt0060161	0.065141	0	0	Beregis Avtomobilya	Innokentiy Smoktunovskiy|Oleg Efremov|Georgi Z...	
10864	21449	tt0061177	0.064317	0	0	What's Up, Tiger Lily?	Tatsuya Mihashi|Akiko Wakabayashi|Mie Hama|Joh...	
10865	22293	tt0060666	0.035919	19000	0	Manos: The Hands of Fate	Harold P. Warren|Tom Neyman|John Reynolds|Dian...	

10866 rows × 21 columns

In [ ]:
```
df.head(5)
```

Out[ ]:

	id	imdb_id	popularity	budget	revenue	original_title	cast	homepage
0	135397	tt0369610	32.985763	150000000	1513528810	Jurassic World	Chris Pratt|Bryce Dallas Howard|Irrfan Khan|Vi...	http://www.jurassicworld.com/
1	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...	http://www.madmaxmovie.com/
2	262500	tt2908446	13.112507	110000000	295238201	Insurgent	Shailene Woodley|Theo James|Kate Winslet|Ansel...	http://www.thedivergentseries.movie/#insurgent
3	140607	tt2488496	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford|Mark Hamill|Carrie Fisher|Adam D...	http://www.starwars.com/films/star-wars-episod...
4	168259	tt2820852	9.335014	190000000	1506249360	Furious 7	Vin Diesel|Paul Walker|Jason Statham|Michelle ...	http://www.furious7.com/

5 rows × 21 columns

In [ ]:
```
df.shape
```

Out[ ]:
```
(10866, 21)
```

In [ ]:
```
df.isnull().sum()
```

Out[ ]:
```
id                        0
imdb_id                  10
popularity                0
budget                    0
revenue                   0
original_title            0
cast                     76
homepage               7930
director                 44
tagline                2824
keywords               1493
overview                  4
runtime                   0
genres                   23
production_companies   1030
release_date              0
vote_count                0
vote_average              0
release_year              0
budget_adj                0
revenue_adj               0
dtype: int64
```

In [ ]:
```
df.columns
```

Out[ ]:
```
Index(['id', 'imdb_id', 'popularity', 'budget', 'revenue', 'original_title',
       'cast', 'homepage', 'director', 'tagline', 'keywords', 'overview',
       'runtime', 'genres', 'production_companies', 'release_date',
       'vote_count', 'vote_average', 'release_year', 'budget_adj',
       'revenue_adj'],
      dtype='object')
```

In [ ]:
```
df.drop(columns=['imdb_id','id','homepage','cast','overview','tagline','budget_adj',],inplace=True)
```

In [ ]:
```
df.head()
```

Out[ ]:

	popularity	budget	revenue	original_title	director	keywords	runtime	genres
0	32.985763	150000000	1513528810	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Action|Adventure|Science Fiction|Thriller
1	28.419936	150000000	378436354	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia	120	Action|Adventure|Science Fiction|Thriller
2	13.112507	110000000	295238201	Insurgent	Robert Schwentke	based on novel|revolution|dystopia|sequel|dyst...	119	Adventure|Science Fiction|Thriller
3	11.173104	200000000	2068178225	Star Wars: The Force Awakens	J.J. Abrams	android|spaceship|jedi|space opera|3d	136	Action|Adventure|Science Fiction|Fantasy
4	9.335014	190000000	1506249360	Furious 7	James Wan	car race|speed|revenge|suspense|car	137	Action|Crime|Thriller

In [ ]:
```
df.isnull().sum()
```

Out[ ]:
```
popularity                0
budget                    0
revenue                   0
original_title            0
director                 44
keywords               1493
runtime                   0
genres                   23
production_companies    1030
release_date              0
vote_count                0
vote_average              0
release_year              0
revenue_adj               0
dtype: int64
```

In [ ]:
```
df.dropna(how='any',subset=['genres',"director"],inplace=True)
```

In [ ]:
```
df.columns
```

Out[ ]:
```
Index(['popularity', 'budget', 'revenue', 'original_title', 'director',
       'keywords', 'runtime', 'genres', 'production_companies', 'release_date',
       'vote_count', 'vote_average', 'release_year', 'revenue_adj'],
      dtype='object')
```

In [ ]:
```
df['production_companies ']=df['production_companies'].fillna(0)
df['keywords']=df['keywords'].fillna(0)
```

In [ ]:
```
df.isnull().sum()
```

Out[ ]:
```
popularity              0
budget                  0
revenue                 0
original_title          0
director                0
keywords                0
runtime                 0
genres                  0
production_companies    0
release_date            0
vote_count              0
vote_average            0
release_year            0
revenue_adj             0
production_companies    0
dtype: int64
```

In [ ]:
```
df.head()
```

Out[ ]:

	popularity	budget	revenue	original_title	director	keywords	runtime	genres
0	32.985763	150000000	1513528810	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Action|Adventure|Science Fiction|Thriller
1	28.419936	150000000	378436354	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia	120	Action|Adventure|Science Fiction|Thriller
2	13.112507	110000000	295238201	Insurgent	Robert Schwentke	based on novel|revolution|dystopia|sequel|dyst...	119	Adventure|Science Fiction|Thriller
3	11.173104	200000000	2068178225	Star Wars: The Force Awakens	J.J. Abrams	android|spaceship|jedi|space opera|3d	136	Action|Adventure|Science Fiction|Fantasy
4	9.335014	190000000	1506249360	Furious 7	James Wan	car race|speed|revenge|suspense|car	137	Action|Crime|Thriller

In [ ]:
```
df['popularity']=df['popularity'].round(2)
```

In [ ]:
```
df.head()
```

Out[ ]:

	popularity	budget	revenue	original_title	director	keywords	runtime	genres
0	32.99	150000000	1513528810	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Action|Adventure|Science Fiction|Thriller
1	28.42	150000000	378436354	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia	120	Action|Adventure|Science Fiction|Thriller
2	13.11	110000000	295238201	Insurgent	Robert Schwentke	based on novel|revolution|dystopia|sequel|dyst...	119	Adventure|Science Fiction|Thriller
3	11.17	200000000	2068178225	Star Wars: The Force Awakens	J.J. Abrams	android|spaceship|jedi|space opera|3d	136	Action|Adventure|Science Fiction|Fantasy
4	9.34	190000000	1506249360	Furious 7	James Wan	car race|speed|revenge|suspense|car	137	Action|Crime|Thriller

In [ ]:
```
df.insert(3,'profit',df.revenue-df.budget)
```

In [ ]:
```
df.insert(4,'ROI',df.profit/df.budget)
```

In [ ]:
```
df
```

Out[ ]:

	popularity	budget	revenue	profit	ROI	original_title	director	keywords
0	32.99	150000000	1513528810	1363528810	9.090192	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island
1	28.42	150000000	378436354	228436354	1.522909	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia
2	13.11	110000000	295238201	185238201	1.683984	Insurgent	Robert Schwentke	based on novel|revolution|dystopia|sequel|dyst...
3	11.17	200000000	2068178225	1868178225	9.340891	Star Wars: The Force Awakens	J.J. Abrams	android|spaceship|jedi|space opera|3d
4	9.34	190000000	1506249360	1316249360	6.927628	Furious 7	James Wan	car race|speed|revenge|suspense|car
...	...	...	...	...	...	...	...	...
10861	0.08	0	0	0	NaN	The Endless Summer	Bruce Brown	surfer|surfboard|surfing
10862	0.07	0	0	0	NaN	Grand Prix	John Frankenheimer	car race|racing|formula 1
10863	0.07	0	0	0	NaN	Beregis Avtomobilya	Eldar Ryazanov	car|trolley|stealing car
10864	0.06	0	0	0	NaN	What's Up, Tiger Lily?	Woody Allen	spoof
10865	0.04	19000	0	-19000	-1.000000	Manos: The Hands of Fate	Harold P. Warren	fire|gun|drive|sacrifice|flashlight

10801 rows × 17 columns

In [ ]:
```
df['profit']=df['profit'].round(2)
df['ROI']=df['ROI'].round(2)
```

In [ ]:
```
df
```

Out[ ]:

	popularity	budget	revenue	profit	ROI	original_title	director	keywords	runti
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	
1	28.42	150000000	378436354	228436354	1.52	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia	
2	13.11	110000000	295238201	185238201	1.68	Insurgent	Robert Schwentke	based on novel|revolution|dystopia|sequel|dyst...	
3	11.17	200000000	2068178225	1868178225	9.34	Star Wars: The Force Awakens	J.J. Abrams	android|spaceship|jedi|space opera|3d	
4	9.34	190000000	1506249360	1316249360	6.93	Furious 7	James Wan	car race|speed|revenge|suspense|car	
...	...	...	...	...	...	...	...	...	
10861	0.08	0	0	0	NaN	The Endless Summer	Bruce Brown	surfer|surfboard|surfing	
10862	0.07	0	0	0	NaN	Grand Prix	John Frankenheimer	car race|racing|formula 1	
10863	0.07	0	0	0	NaN	Beregis Avtomobilya	Eldar Ryazanov	car|trolley|stealing car	
10864	0.06	0	0	0	NaN	What's Up, Tiger Lily?	Woody Allen	spoof	
10865	0.04	19000	0	-19000	-1.00	Manos: The Hands of Fate	Harold P. Warren	fire|gun|drive|sacrifice|flashlight	

10801 rows × 17 columns

In [ ]:
```
df1=[['popularity','budget','revenue','profit','ROI','vote_count','vote_average','release_year']]
```

In [ ]:
```
df.isnull().sum()
```

Out[ ]:
```
popularity                0
budget                    0
revenue                   0
profit                    0
ROI                    4641
original_title            0
director                  0
keywords                  0
runtime                   0
genres                    0
production_companies      0
release_date              0
vote_count                0
vote_average              0
release_year              0
revenue_adj               0
production_companies      0
dtype: int64
```

In [ ]:
```
non_finite_value= ~np.isfinite(df['ROI'])
```

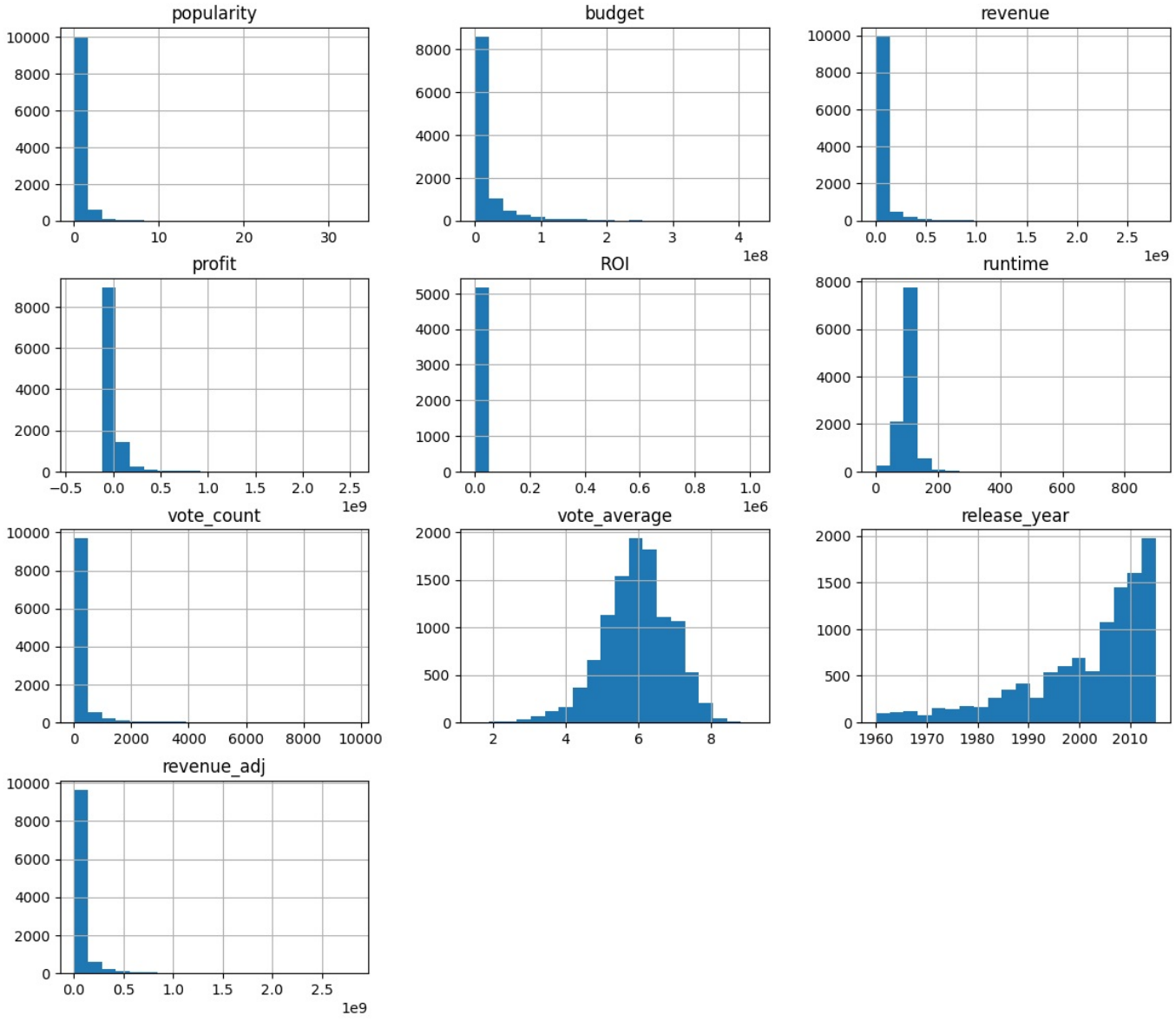
In [ ]:
```
non_finite_value.sum()
```

Out[ ]:
```
5636
```

In [ ]:
```
df['ROI']=df['ROI'].replace([np.inf,-np.inf],np.nan)
```

In [ ]:
```
df1=[['popularity','budget','revenue','profit','ROI','vote_count','vote_average','release_year']]
```

In [ ]:
```
df.hist(bins=20,figsize=(14,12))
plt.show()
```



In [ ]:
```
df.popularity.value_counts()
```

Out[ ]:
```
popularity
0.14    193
0.28    190
0.21    186
0.25    182
0.20    179
       ...
5.81      1
5.08      1
3.83      1
3.74      1
2.68      1
Name: count, Length: 483, dtype: int64
```

In [ ]:
```
df.head(5)
```

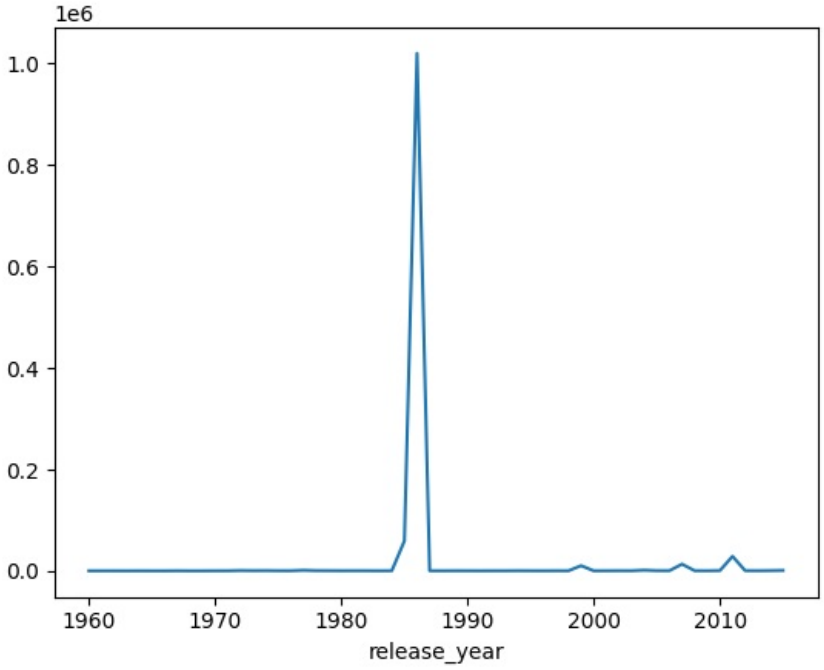
Out[ ]:

	popularity	budget	revenue	profit	ROI	original_title	director	keywords	runtime	
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Acti
1	28.42	150000000	378436354	228436354	1.52	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia	120	Acti
2	13.11	110000000	295238201	185238201	1.68	Insurgent	Robert Schwentke	based on novel|revolution|dystopia|sequel|dyst...	119	
3	11.17	200000000	2068178225	1868178225	9.34	Star Wars: The Force Awakens	J.J. Abrams	android|spaceship|jedi|space opera|3d	136	Acti
4	9.34	190000000	1506249360	1316249360	6.93	Furious 7	James Wan	car race|speed|revenge|suspense|car	137	

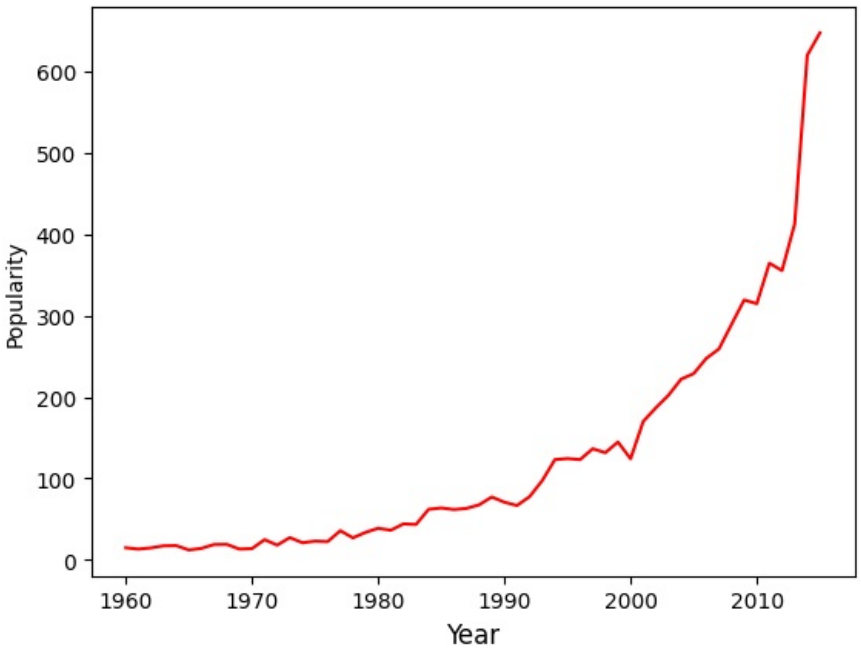
In [ ]:
```
df2=df.groupby('release_year')['ROI'].sum()
```

In [ ]:
```
df2.plot(kind='line')
```

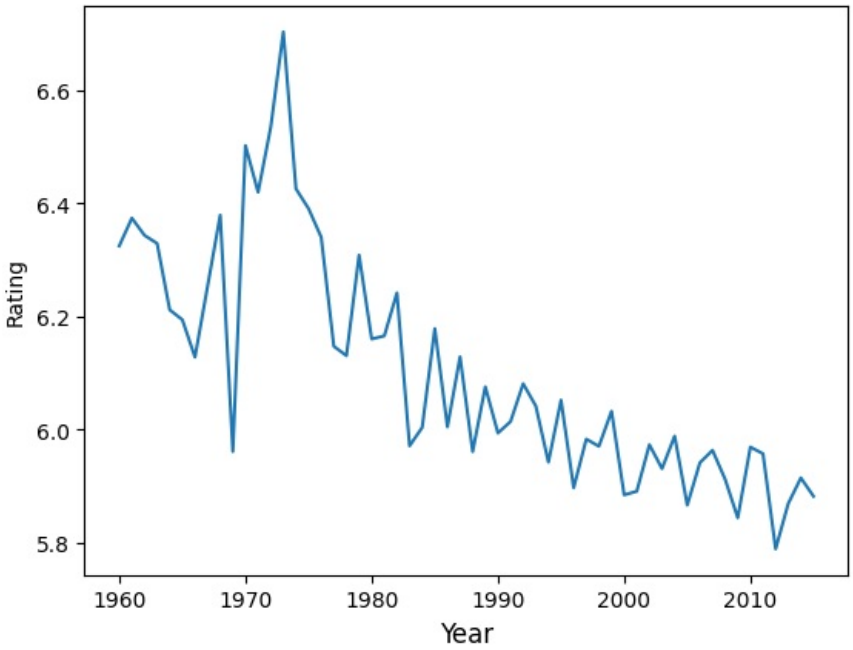
Out[ ]:
```
<Axes: xlabel='release_year'>
```



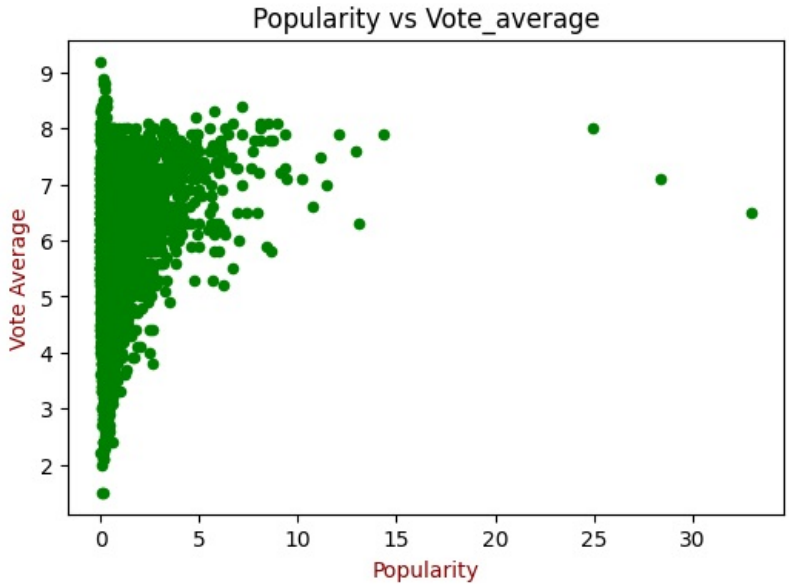
In [ ]:
```
df3=df.groupby('release_year')['popularity'].sum().plot(kind='line',color='red')
plt.xlabel('Year',fontsize=12)
plt.ylabel('Popularity')
plt.show()
```



In [ ]:
```
df4=df.groupby('release_year')['vote_average'].mean()
df4.plot(kind='line')
plt.xlabel('Year',fontsize=12)
plt.ylabel('Rating')
plt.show()
```



In [ ]:
```
df5= df.plot.scatter(x='popularity',y='vote_average',color= 'green',figsize=(6,4))
df5.set_xlabel('Popularity',color='Darkred')
df5.set_ylabel('Vote Average',color='Darkred')
df5.set_title('Popularity vs Vote_average',fontsize=12)
plt.show()
```



In [ ]:
```
df.genres.value_counts()
```

Out[ ]:
```
genres
Drama                                          711
Comedy                                         707
Documentary                                    306
Drama|Romance                                  289
Comedy|Drama                                   280
                                              ...
Science Fiction|Horror|Action|Thriller           1
Action|Thriller|Science Fiction|Mystery          1
Comedy|Music|Romance|Foreign                     1
Documentary|Drama|Comedy                         1
Mystery|Science Fiction|Thriller|Drama           1
Name: count, Length: 2031, dtype: int64
```

In [ ]:
```
split=['genres']
for i in split :
    df[i]=df[i].apply(lambda x : x.split("|"))
df.head(3)
```

Out[ ]:

	popularity	budget	revenue	profit	ROI	original_title	director	keywords	runtime	
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Adv S
1	28.42	150000000	378436354	228436354	1.52	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia	120	Adv S
2	13.11	110000000	295238201	185238201	1.68	Insurgent	Robert Schwentke	based on novel|revolution|dystopia|sequel|dyst...	119	[Adv S

In [ ]:
```
df = df.explode('genres')
df
```

Out[ ]:

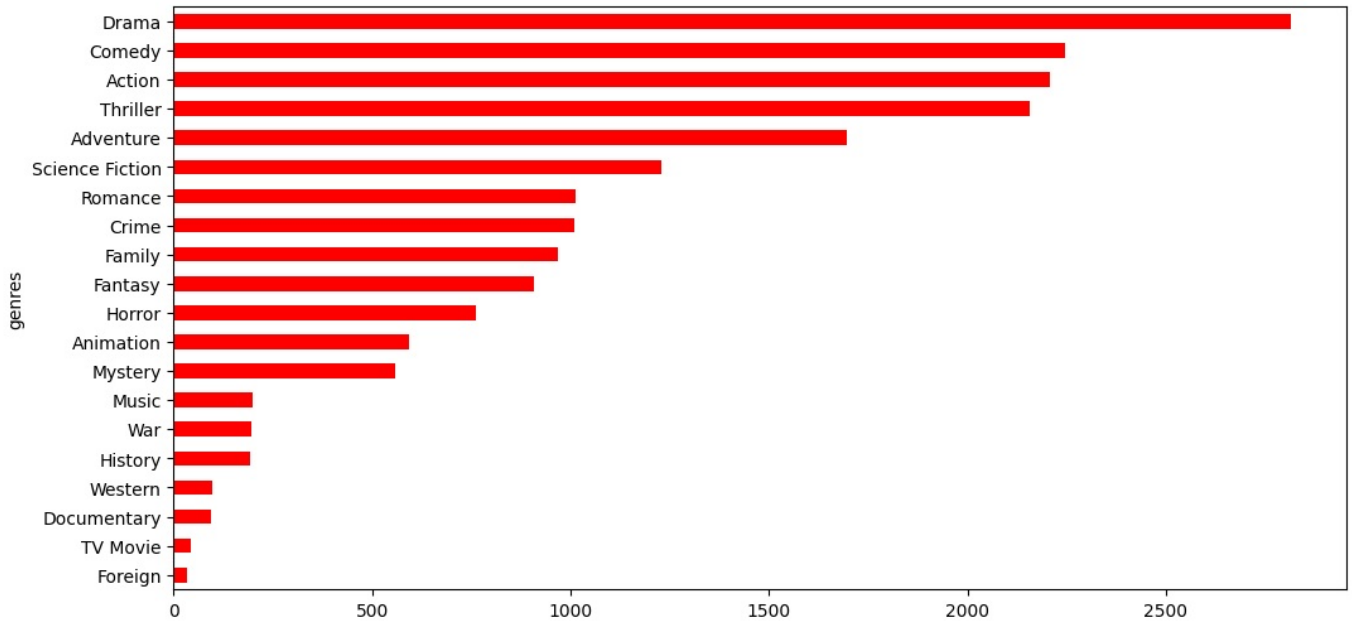
	popularity	budget	revenue	profit	ROI	original_title	director	keywords	runtime	gen
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Ac
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Adven
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Scie Fic
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Thr
1	28.42	150000000	378436354	228436354	1.52	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia	120	Ac
...	...	...	...	...	...	...	...	...	...	
10863	0.07	0	0	0	NaN	Beregis Avtomobilya	Eldar Ryazanov	car|trolley|stealing car	94	Mys
10863	0.07	0	0	0	NaN	Beregis Avtomobilya	Eldar Ryazanov	car|trolley|stealing car	94	Com
10864	0.06	0	0	0	NaN	What's Up, Tiger Lily?	Woody Allen	spoof	80	Ac
10864	0.06	0	0	0	NaN	What's Up, Tiger Lily?	Woody Allen	spoof	80	Com
10865	0.04	19000	0	-19000	-1.00	Manos: The Hands of Fate	Harold P. Warren	fire|gun|drive|sacrifice|flashlight	74	Ho

26869 rows × 17 columns

In [ ]:
```
df7=df.groupby('genres')['popularity'].sum().sort_values(ascending=True)
df7
```

Out[ ]:
```
genres
Foreign              35.24
TV Movie             44.03
Documentary          93.13
Western              97.42
History             192.35
War                 196.48
Music               198.15
Mystery             558.55
Animation           594.46
Horror              761.39
Fantasy             908.87
Family              967.06
Crime              1009.07
Romance            1013.21
Science Fiction    1230.41
Adventure          1697.11
Thriller           2155.90
Action             2208.08
Comedy             2246.25
Drama              2815.43
Name: popularity, dtype: float64
```

In [ ]:
```
df7.plot(kind='barh',x='genres',y= 'popularity',color='red',figsize=(12,6))
plt.show()
```



In [ ]:
```
df.head(1)
```

Out[ ]:

	popularity	budget	revenue	profit	ROI	original_title	director	keywords	runtime	genres	produ
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Action	Univer Enterta

In [ ]:
```
df.dtypes
```

Out[ ]:
```
popularity               float64
budget                     int64
revenue                    int64
profit                     int64
ROI                      float64
original_title            object
director                  object
keywords                  object
runtime                    int64
genres                    object
production_companies      object
release_date              object
vote_count                 int64
vote_average             float64
release_year               int64
revenue_adj              float64
production_companies      object
dtype: object
```

In [ ]:
```
df['release_date']=pd.to_datetime(df['release_date'])
```

```
C:\Users\APURVA BOBADE\AppData\Local\Temp\ipykernel_15384\642638228.py:1: UserWarning: Could not infer format, s
o each element will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-e
xpected, please specify a format.
  df['release_date']=pd.to_datetime(df['release_date'])
```

In [ ]:
```
df.dtypes
```

Out[ ]:
```
popularity                      float64
budget                            int64
revenue                           int64
profit                            int64
ROI                             float64
original_title                   object
director                         object
keywords                         object
runtime                           int64
genres                           object
production_companies             object
release_date              datetime64[ns]
vote_count                        int64
vote_average                    float64
release_year                      int64
revenue_adj                     float64
production_companies             object
dtype: object
```

In [ ]:
```
df['extracted_month']= df['release_date'].dt.month
```

In [ ]:
```
df.head()
```

Out[ ]:

	popularity	budget	revenue	profit	ROI	original_title	director	keywords	runtime	genres	p
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Action	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Adventure	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Science Fiction	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Thriller	U E
1	28.42	150000000	378436354	228436354	1.52	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia	120	Action	

In [ ]:
```
df8=df.groupby('extracted_month')['popularity'].sum()
```

In [ ]:
```
df8
```

Out[ ]:
```
extracted_month
1     1131.78
2     1092.93
3     1458.32
4     1191.81
5     1687.53
6     1936.84
7     1694.03
8     1432.59
9     1872.28
10    1811.91
11    1710.35
12    2002.22
Name: popularity, dtype: float64
```

In [ ]:
```
df8.index
```

Out[ ]:
```
Index([1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12], dtype='int32', name='extracted_month')
```

In [ ]:
```
df8.values
```

Out[ ]:
```
array([1131.78, 1092.93, 1458.32, 1191.81, 1687.53, 1936.84, 1694.03,
       1432.59, 1872.28, 1811.91, 1710.35, 2002.22])
```

In [ ]:
```
data={'extracted_month':df8.index,'popularity':df8.values


}
df8=pd.DataFrame(data)
```

In [ ]:
```
df8
```

Out[ ]:

	extracted_month	popularity
0	1	1131.78
1	2	1092.93
2	3	1458.32
3	4	1191.81
4	5	1687.53
5	6	1936.84
6	7	1694.03
7	8	1432.59
8	9	1872.28
9	10	1811.91
10	11	1710.35
11	12	2002.22

In [ ]:
```
index_to_month={
    1:'Jan',2:'Feb',3:'Mar',4:'Apr',5:'May',6:'Jun',7:'Jul',8:'Aug',9:'Sept',
    10:'Oct',11:'Nov',12:'Dec'
```

```
}
```

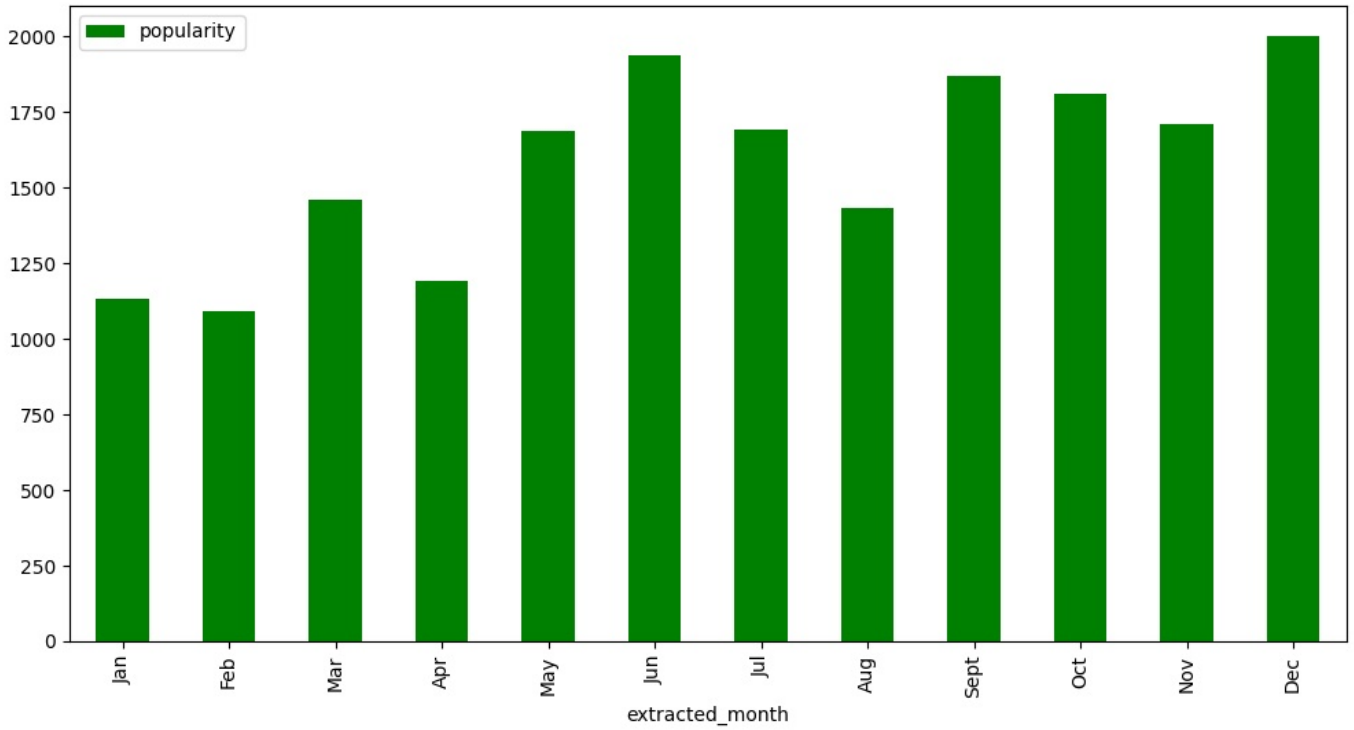
In [ ]:
```
df8.extracted_month=df8.extracted_month.map(index_to_month)
```

In [ ]:
```
df8
```

Out[ ]:

	extracted_month	popularity
0	Jan	1131.78
1	Feb	1092.93
2	Mar	1458.32
3	Apr	1191.81
4	May	1687.53
5	Jun	1936.84
6	Jul	1694.03
7	Aug	1432.59
8	Sept	1872.28
9	Oct	1811.91
10	Nov	1710.35
11	Dec	2002.22

In [ ]:
```
df8.plot(kind='bar',x='extracted_month',y='popularity',color='green',figsize=(12,6))
plt.show()
```



In [ ]:
```
df9 = df.groupby('extracted_month')['revenue'].sum()
df9
```

Out[ ]:
```
extracted_month
1       35873456579
2       54352852344
3       93669046441
4       77813179749
5      151475532493
6      193681776686
7      141947570995
8       71642408883
9       70379641581
10      84054172048
11     139176268899
12     164738399960
Name: revenue, dtype: int64
```

In [ ]:
```
data={'extracted_month':df9.index,'revenue':df9.values


}
```

```
df9=pd.DataFrame(data)
```

In [ ]:
```
index_to_month={
    1:'Jan',2:'Feb',3:'Mar',4:'Apr',5:'May',6:'Jun',7:'Jul',8:'Aug',9:'Sept',
    10:'Oct',11:'Nov',12:'Dec'
}
```

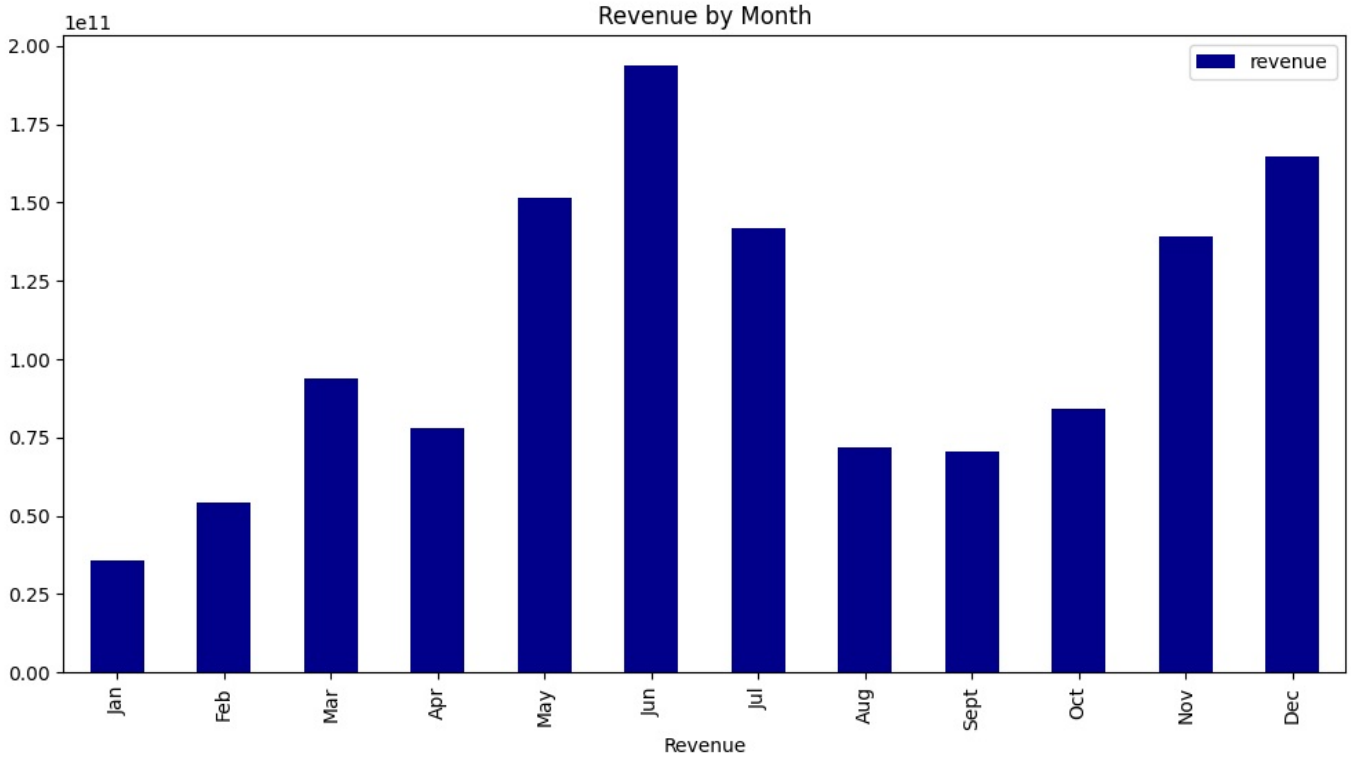
In [ ]:
```
df9.extracted_month=df8.extracted_month.map(index_to_month)
```

In [ ]:
```
df9
```

Out[ ]:

	extracted_month	revenue
0	Jan	35873456579
1	Feb	54352852344
2	Mar	93669046441
3	Apr	77813179749
4	May	151475532493
5	Jun	193681776686
6	Jul	141947570995
7	Aug	71642408883
8	Sept	70379641581
9	Oct	84054172048
10	Nov	139176268899
11	Dec	164738399960

In [ ]:
```
df9.plot(kind='bar',x='extracted_month',y='revenue',color='darkblue',figsize=(12,6))
plt.title('Revenue by Month')
plt.xlabel("Revenue")
plt.ylabel=("Month")
plt.show()
```



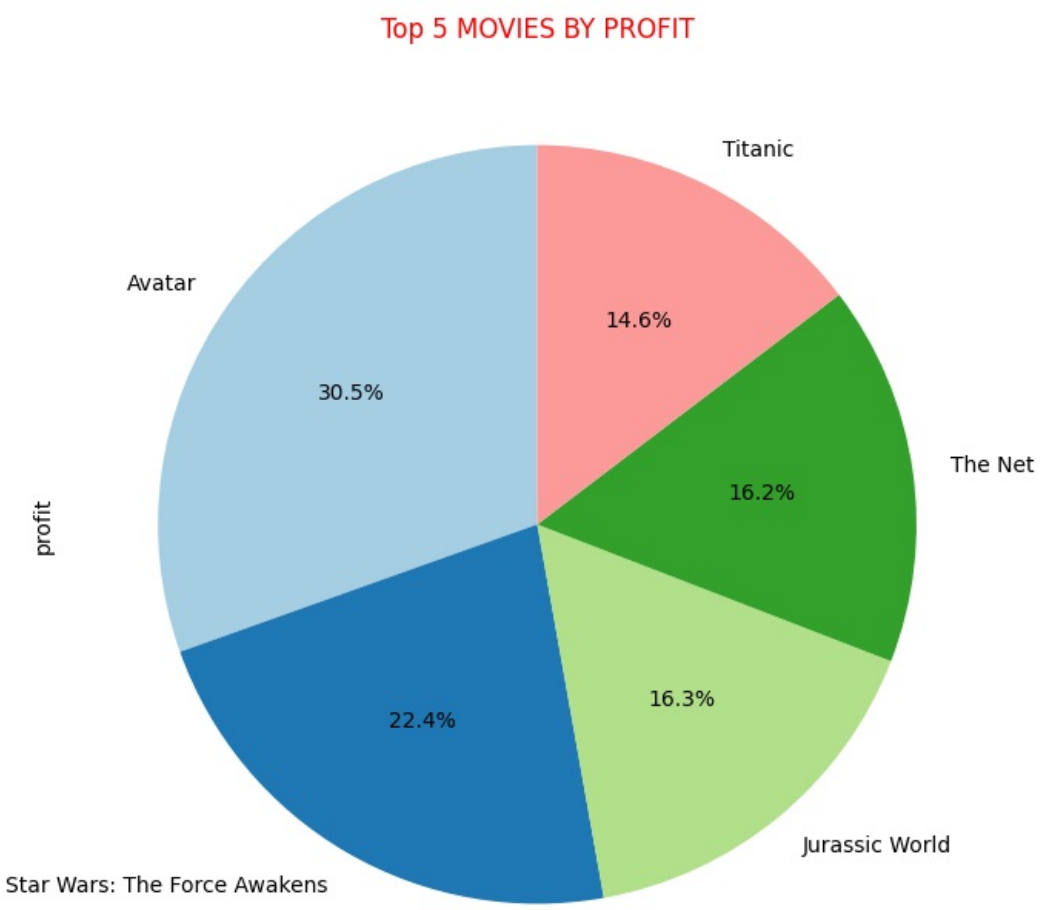
In [ ]:
```
df.head()
```

Out[ ]:

	popularity	budget	revenue	profit	ROI	original_title	director	keywords	runtime	genres	p
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Action	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Adventure	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Science Fiction	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Thriller	U E
1	28.42	150000000	378436354	228436354	1.52	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia	120	Action	

In [ ]:
```
df10=df.groupby('original_title')['profit'].sum().sort_values(ascending=False).head(5)
```

In [ ]:
```
df10.plot(kind='pie',autopct='%1.1f%%',startangle=90,figsize=(14,8),colors = plt.cm.Paired.colors)
plt.title('Top 5 MOVIES BY PROFIT',color = 'red')
plt.show()
```



In [ ]:
```
df.head()
```

Out[ ]:

	popularity	budget	revenue	profit	ROI	original_title	director	keywords	runtime	genres	p
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Action	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Adventure	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Science Fiction	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Thriller	U E
1	28.42	150000000	378436354	228436354	1.52	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia	120	Action	

In [ ]:
```
df11=df.production_companies.value_counts().head(5)
df11
```

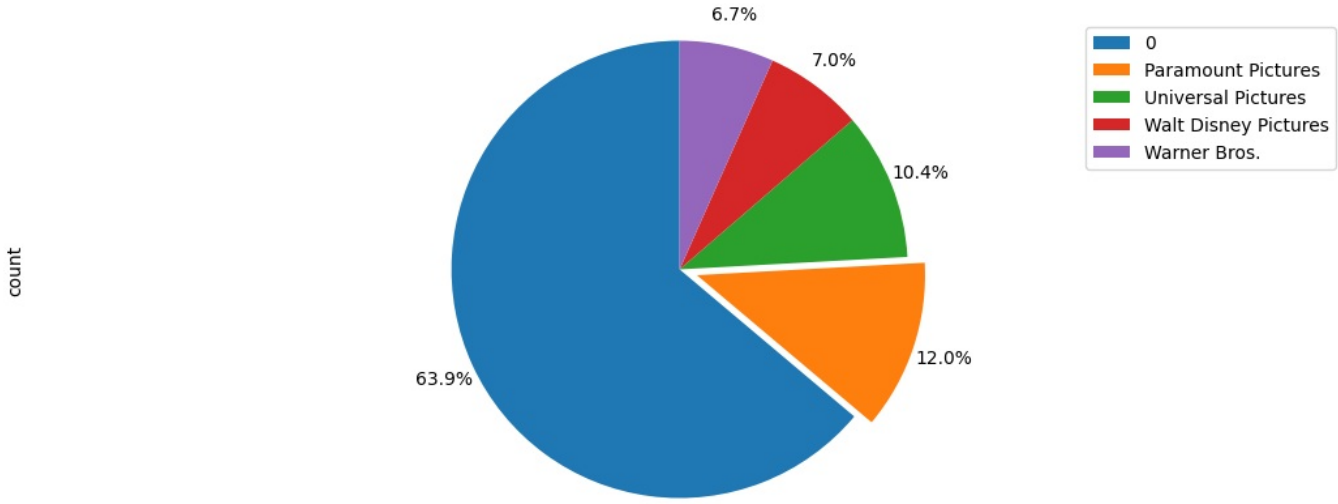
Out[ ]:
```
production_companies
0                       2152
Paramount Pictures       404
Universal Pictures       352
Walt Disney Pictures     236
Warner Bros.             225
Name: count, dtype: int64
```

In [ ]:
```
df11.index
```

Out[ ]:
```
Index([                    0,   'Paramount Pictures',   'Universal Pictures',
       'Walt Disney Pictures',         'Warner Bros.'],
      dtype='object', name='production_companies')
```

In [ ]:
```
explode_list= [0,0.08,0,0,0]
df11.plot(kind='pie',autopct='%1.1f%%',startangle=90,figsize=(13,5),labels=None,pctdistance=1.14,
          explode=explode_list)

#plt.title('Top 5 MOVIES BY PROFIT',color = 'red')
plt.legend(labels=df11.index,loc='upper right')
plt.axis('equal')
plt.show()
```



In [ ]:
```
df.head()
```

Out[ ]:

	popularity	budget	revenue	profit	ROI	original_title	director	keywords	runtime	genres	p
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Action	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Adventure	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Science Fiction	U E
0	32.99	150000000	1513528810	1363528810	9.09	Jurassic World	Colin Trevorrow	monster|dna|tyrannosaurus rex|velociraptor|island	124	Thriller	U E
1	28.42	150000000	378436354	228436354	1.52	Mad Max: Fury Road	George Miller	future|chase|post-apocalyptic|dystopia|australia	120	Action	

In [ ]:
```
df12=df.keywords.value_counts().head(15)
df12
```

Out[ ]:
```
keywords
0                                  3262
woman director                      238
independent film                    177
suspense                             68
musical                              57
sport                                50
duringcreditsstinger                 49
holiday                              38
biography                            26
dystopia                             26
independent film|woman director      23
holiday|christmas                    23
sequel                               19
based on novel                       18
christmas                            18
Name: count, dtype: int64
```

In [ ]:
```
df12.index
```

Out[ ]:
```
Index([                              0,                  'woman director',
                      'independent film',                        'suspense',
                               'musical',                          'sport',
                   'duringcreditsstinger',                        'holiday',
                               'biography',                        'dystopia',
       'independent film|woman director',               'holiday|christmas',
                                'sequel',                  'based on novel',
                              'christmas'],
      dtype='object', name='keywords')
```

In [ ]:
```
df12.values
```

Out[ ]:
```
array([3262,  238,  177,   68,   57,   50,   49,   38,   26,   26,   23,
         23,   19,   18,   18], dtype=int64)
```

In [ ]:
```
data={'keywords':df12.index,'value':df12.values


}
df12=pd.DataFrame(data)
```

In [ ]:
```
df12
```

Out[ ]:

	keywords	value
0	0	3262
1	woman director	238
2	independent film	177
3	suspense	68
4	musical	57
5	sport	50
6	duringcreditsstinger	49
7	holiday	38
8	biography	26
9	dystopia	26
10	independent film|woman director	23
11	holiday|christmas	23
12	sequel	19
13	based on novel	18
14	christmas	18

In [ ]:

In [ ]: