In [53]:
```
import numpy as np
import pandas as pd
import warnings
```

In [54]:
```
movies=pd.read_csv("movie.csv" ,sep=',')
```

In [55]:
```
movies
```

Out[55]:

	movieId	title	genres
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
1	2	Jumanji (1995)	Adventure|Children|Fantasy
2	3	Grumpier Old Men (1995)	Comedy|Romance
3	4	Waiting to Exhale (1995)	Comedy|Drama|Romance
4	5	Father of the Bride Part II (1995)	Comedy
...	...	...	...
27273	131254	Kein Bund für's Leben (2007)	Comedy
27274	131256	Feuer, Eis & Dosenbier (2002)	Comedy
27275	131258	The Pirates (2014)	Adventure
27276	131260	Rentun Ruusu (2001)	(no genres listed)
27277	131262	Innocence (2014)	Adventure|Fantasy|Horror

27278 rows × 3 columns

In [56]:
```
movies.head(12)
```

Out[56]:

	movieId	title	genres
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
1	2	Jumanji (1995)	Adventure|Children|Fantasy
2	3	Grumpier Old Men (1995)	Comedy|Romance
3	4	Waiting to Exhale (1995)	Comedy|Drama|Romance
4	5	Father of the Bride Part II (1995)	Comedy
5	6	Heat (1995)	Action|Crime|Thriller
6	7	Sabrina (1995)	Comedy|Romance
7	8	Tom and Huck (1995)	Adventure|Children
8	9	Sudden Death (1995)	Action
9	10	GoldenEye (1995)	Action|Adventure|Thriller
10	11	American President, The (1995)	Comedy|Drama|Romance
11	12	Dracula: Dead and Loving It (1995)	Comedy|Horror

In [57]:
```
ratings=pd.read_csv("rating.csv" ,sep=',')
```

In [58]:
```
ratings
```

Out[58]:

	userId	movieId	rating	timestamp
0	1	2	3.5	2005-04-02 23:53:47
1	1	29	3.5	2005-04-02 23:31:16
2	1	32	3.5	2005-04-02 23:33:39
3	1	47	3.5	2005-04-02 23:32:07
4	1	50	3.5	2005-04-02 23:29:40
...	...	...	...	...
20000258	138493	68954	4.5	2009-11-13 15:42:00
20000259	138493	69526	4.5	2009-12-03 18:31:48
20000260	138493	69644	3.0	2009-12-07 18:10:57
20000261	138493	70286	5.0	2009-11-13 15:42:24
20000262	138493	71619	2.5	2009-10-17 20:25:36

20000263 rows × 4 columns

In [59]:
```
ratings.head(10)
```

Out[59]:

	userId	movieId	rating	timestamp
0	1	2	3.5	2005-04-02 23:53:47
1	1	29	3.5	2005-04-02 23:31:16
2	1	32	3.5	2005-04-02 23:33:39
3	1	47	3.5	2005-04-02 23:32:07
4	1	50	3.5	2005-04-02 23:29:40
5	1	112	3.5	2004-09-10 03:09:00
6	1	151	4.0	2004-09-10 03:08:54
7	1	223	4.0	2005-04-02 23:46:13
8	1	253	4.0	2005-04-02 23:35:40
9	1	260	4.0	2005-04-02 23:33:46

In [60]:
```
tags=pd.read_csv("tag.csv", sep=',')
```

In [61]:
```
tags
```

Out[61]:

	userId	movieId	tag	timestamp
0	18	4141	Mark Waters	2009-04-24 18:19:40
1	65	208	dark hero	2013-05-10 01:41:18
2	65	353	dark hero	2013-05-10 01:41:19
3	65	521	noir thriller	2013-05-10 01:39:43
4	65	592	dark hero	2013-05-10 01:41:18
...	...	...	...	...
465559	138446	55999	dragged	2013-01-23 23:29:32
465560	138446	55999	Jason Bateman	2013-01-23 23:29:38
465561	138446	55999	quirky	2013-01-23 23:29:38
465562	138446	55999	sad	2013-01-23 23:29:32
465563	138472	923	rise to power	2007-11-02 21:12:47

465564 rows × 4 columns

In [62]:
```
tags.head(10)
```

Out[62]:

	userId	movieId	tag	timestamp
0	18	4141	Mark Waters	2009-04-24 18:19:40
1	65	208	dark hero	2013-05-10 01:41:18
2	65	353	dark hero	2013-05-10 01:41:19
3	65	521	noir thriller	2013-05-10 01:39:43
4	65	592	dark hero	2013-05-10 01:41:18
5	65	668	bollywood	2013-05-10 01:37:56
6	65	898	screwball comedy	2013-05-10 01:42:40
7	65	1248	noir thriller	2013-05-10 01:39:43
8	65	1391	mars	2013-05-10 01:40:55
9	65	1617	neo-noir	2013-05-10 01:43:37

In [63]:
```
del ratings['timestamp']
```

In [64]:
```
del tags['timestamp']
```

In [66]:
```
row_0 = tags.iloc[0]
type(row_0)
```

Out[66]:
```
pandas.core.series.Series
```

In [67]:
```
print(row_0)
```

```
userId              18
movieId           4141
tag        Mark Waters
Name: 0, dtype: object
```

In [68]:
```
row_0.index
```

Out[68]:
```
Index(['userId', 'movieId', 'tag'], dtype='object')
```

In [ ]:
```

```

In [71]:
```
row_0['userId']
```

Out[71]:
```
18
```

In [76]:
```
'rating' in row_0
```

Out[76]:
```
False
```

In [77]:
```
row_0.name
```

Out[77]:
```
0
```

In [78]:
```
row_0=row_0.rename('firstrow')
```

In [80]:
```
row_0.name
```

Out[80]:
```
'firstrow'
```

data frames

In [81]:
```
tags.head()
```

Out[81]:

	userId	movieId	tag
0	18	4141	Mark Waters
1	65	208	dark hero
2	65	353	dark hero
3	65	521	noir thriller
4	65	592	dark hero

In [82]:
```
tags.index
```

Out[82]:
```
RangeIndex(start=0, stop=465564, step=1)
```

In [83]:
```
tags.columns
```

Out[83]:
```
Index(['userId', 'movieId', 'tag'], dtype='object')
```

In [89]:
```
tags.iloc[[0,11,500]]
```

Out[89]:

	userId	movieId	tag
0	18	4141	Mark Waters
11	65	1783	noir thriller
500	342	55908	entirely dialogue

In [91]:
```
ratings['rating'].describe()
```

Out[91]:
```
count    2.000026e+07
mean     3.525529e+00
std      1.051989e+00
min      5.000000e-01
25%      3.000000e+00
50%      3.500000e+00
75%      4.000000e+00
max      5.000000e+00
Name: rating, dtype: float64
```

In [92]:
```
rating.describe()
```

Out[92]:

	userId	movieId	rating
count	2.000026e+07	2.000026e+07	2.000026e+07
mean	6.904587e+04	9.041567e+03	3.525529e+00
std	4.003863e+04	1.978948e+04	1.051989e+00
min	1.000000e+00	1.000000e+00	5.000000e-01
25%	3.439500e+04	9.020000e+02	3.000000e+00
50%	6.914100e+04	2.167000e+03	3.500000e+00
75%	1.036370e+05	4.770000e+03	4.000000e+00
max	1.384930e+05	1.312620e+05	5.000000e+00

In [93]:
```
ratings['rating'].mean()
```

Out[93]:
```
3.5255285642993797
```

In [95]:
```
ratings.mean()
```

Out[95]:
```
userId     69045.872583
movieId     9041.567330
rating         3.525529
dtype: float64
```

In [96]:
```
ratings.corr()
```

Out[96]:

	userId	movieId	rating
userId	1.000000	-0.000850	0.001175
movieId	-0.000850	1.000000	0.002606
rating	0.001175	0.002606	1.000000

In [97]:
```
filter1=rating['rating']>10
```

In [98]:
```
filter1
```

Out[98]:
```
0           False
1           False
2           False
3           False
4           False
            ...
20000258    False
20000259    False
20000260    False
20000261    False
20000262    False
Name: rating, Length: 20000263, dtype: bool
```

In [99]:
```
filter1.any()
```

Out[99]:
```
False
```

In [103…
```
filter2=ratings['rating']>0
filter2.all()
```

Out[103]:
```
True
```

data cleanings and handelling missing data

In [105…
```
movies.shape
```

Out[105]:
```
(27278, 3)
```

In [106…
```
movies.isnull()
```

Out[106]:

	movieId	title	genres
0	False	False	False
1	False	False	False
2	False	False	False
3	False	False	False
4	False	False	False
...	...	...	...
27273	False	False	False
27274	False	False	False
27275	False	False	False
27276	False	False	False
27277	False	False	False

27278 rows × 3 columns

In [107…
```
movies.isnull().count()
```

Out[107]:
```
movieId    27278
title      27278
genres     27278
dtype: int64
```

In [108…
```
movies.isnull().any()
```

Out[108]:
```
movieId    False
title      False
genres     False
dtype: bool
```

In [109…
```
movies.isnull().any().any()
```

Out[109]:
```
False
```

In [110…
```
ratings.shape
```

Out[110]:
```
(20000263, 3)
```

In [111…
```
rating.isnull()
```

Out[111]:

	userId	movieId	rating
0	False	False	False
1	False	False	False
2	False	False	False
3	False	False	False
4	False	False	False
...	...	...	...
20000258	False	False	False
20000259	False	False	False
20000260	False	False	False
20000261	False	False	False
20000262	False	False	False

20000263 rows × 3 columns

In [112…
```
rating.isnull().any()
```

Out[112]:
```
userId     False
movieId    False
rating     False
dtype: bool
```

In [113…
```
rating.isnull().any().any()
```

Out[113]:
```
False
```

In [115…
```
tags.shape
```

Out[115]:
```
(465564, 3)
```

In [117…
```
tags.isnull()
```

Out[117]:

	userId	movieId	tag
0	False	False	False
1	False	False	False
2	False	False	False
3	False	False	False
4	False	False	False
...	...	...	...
465559	False	False	False
465560	False	False	False
465561	False	False	False
465562	False	False	False
465563	False	False	False

465564 rows × 3 columns

In [118...
```
tags.isnull().sum()
```

Out[118]:
```
userId      0
movieId     0
tag        16
dtype: int64
```

In [119...
```
tags.isnull().any().any()
```

Out[119]:
```
True
```

In [121...
```
tags=tags.dropna()
```

In [124...
```
tags.isnull().any().any()
```

Out[124]:
```
False
```

In [126...
```
tags.shape
```

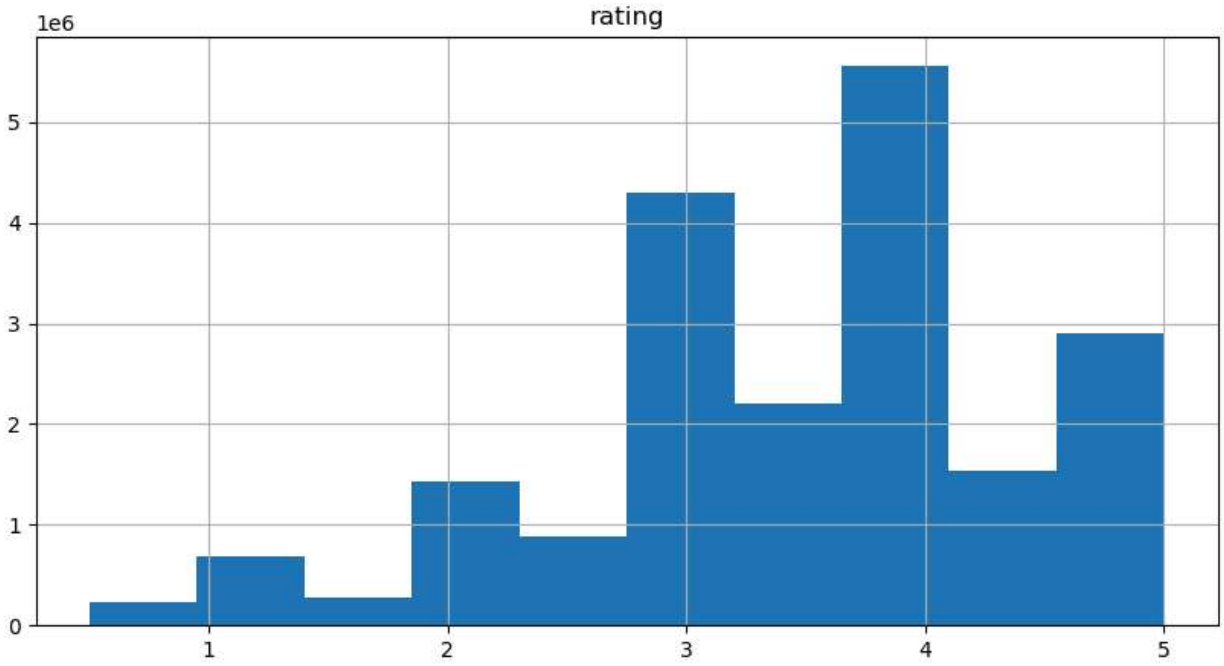
Out[126]:
```
(465548, 3)
```

data visulizations

In [127...
```
%matplotlib inline
```

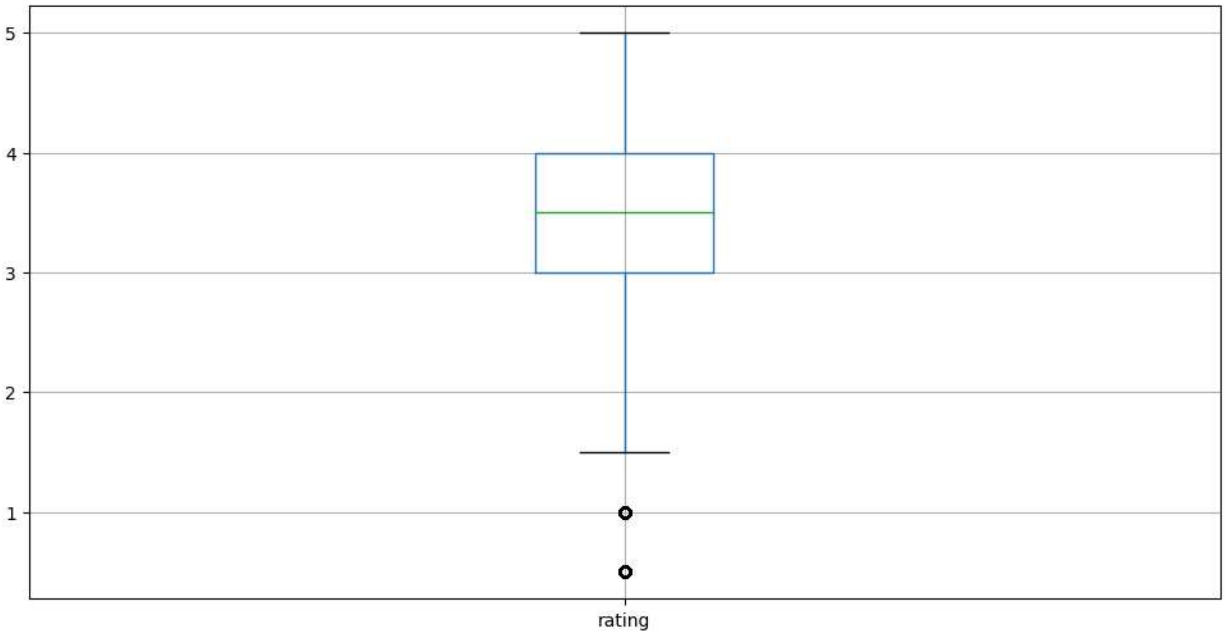
In [128...
```
ratings.hist(column='rating',figsize=(10,5))
```

Out[128]:
```
array([[<Axes: title={'center': 'rating'}>]], dtype=object)
```



In [129…
```
ratings.boxplot(column='rating',figsize=(12,6))
```

Out[129]:
```
<Axes: >
```



slicing out columns

In [130…
```
tags['tag'].head()
```

Out[130]:
```
0      Mark Waters
1        dark hero
2        dark hero
3    noir thriller
4        dark hero
Name: tag, dtype: object
```

In [131…
```
movies[['title','genres']].head()
```

Out[131]:

	title	genres
0	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
1	Jumanji (1995)	Adventure|Children|Fantasy
2	Grumpier Old Men (1995)	Comedy|Romance
3	Waiting to Exhale (1995)	Comedy|Drama|Romance
4	Father of the Bride Part II (1995)	Comedy

In [135…
```
ratings[-10:]
```

Out[135]:

	userId	movieId	rating
20000253	138493	60816	4.5
20000254	138493	61160	4.0
20000255	138493	65682	4.5
20000256	138493	66762	4.5
20000257	138493	68319	4.5
20000258	138493	68954	4.5
20000259	138493	69526	4.5
20000260	138493	69644	3.0
20000261	138493	70286	5.0
20000262	138493	71619	2.5

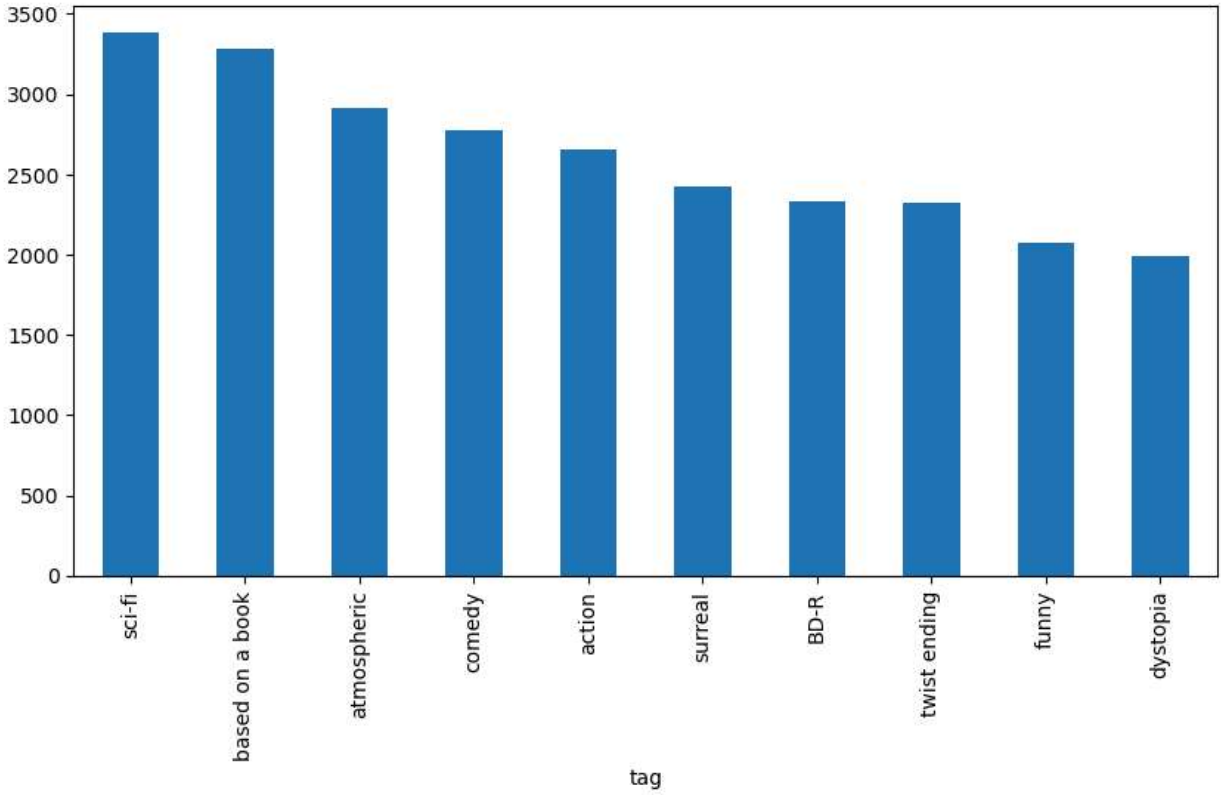
In [136…
```
tag_counts=tags['tag'].value_counts()
```

In [137…
```
tag_counts[-10:]
```

Out[137]:
```
tag
missing child                  1
Ron Moore                      1
Citizen Kane                   1
mullet                         1
biker gang                     1
Paul Adelstein                 1
the wig                        1
killer fish                    1
genetically modified monsters  1
topless scene                  1
Name: count, dtype: int64
```

In [139…
```
tag_counts[:10].plot(kind='bar',figsize=(10,5))
```

Out[139]:
```
<Axes: xlabel='tag'>
```



filtering and sellescing rows

In [141…
```
is_highly_rated=ratings['rating']>=5.0
```

In [144…
```
ratings[is_highly_rated][30:60]
```

Out[144]:

	userId	movieId	rating
239	3	50	5.0
242	3	175	5.0
244	3	223	5.0
245	3	260	5.0
246	3	316	5.0
247	3	318	5.0
248	3	329	5.0
252	3	457	5.0
253	3	480	5.0
254	3	490	5.0
256	3	541	5.0
258	3	593	5.0
263	3	858	5.0
264	3	904	5.0
267	3	924	5.0
268	3	953	5.0
271	3	1060	5.0
272	3	1073	5.0
275	3	1084	5.0
276	3	1089	5.0
278	3	1097	5.0
282	3	1129	5.0
286	3	1196	5.0
287	3	1197	5.0
288	3	1198	5.0
291	3	1206	5.0
292	3	1208	5.0
293	3	1210	5.0
294	3	1213	5.0
295	3	1214	5.0

In [149…

```
is_action=movies['genres'].str.contains('Action')
movies[is_action][0:20]
```

Out[149]:

	movieId	title	genres
5	6	Heat (1995)	Action|Crime|Thriller
8	9	Sudden Death (1995)	Action
9	10	GoldenEye (1995)	Action|Adventure|Thriller
14	15	Cutthroat Island (1995)	Action|Adventure|Romance
19	20	Money Train (1995)	Action|Comedy|Crime|Drama|Thriller
22	23	Assassins (1995)	Action|Crime|Thriller
41	42	Dead Presidents (1995)	Action|Crime|Drama
43	44	Mortal Kombat (1995)	Action|Adventure|Fantasy
50	51	Guardian Angel (1994)	Action|Drama|Thriller
65	66	Lawnmower Man 2: Beyond Cyberspace (1996)	Action|Sci-Fi|Thriller
69	70	From Dusk Till Dawn (1996)	Action|Comedy|Horror|Thriller
70	71	Fair Game (1995)	Action
75	76	Screamers (1995)	Action|Sci-Fi|Thriller
77	78	Crossing Guard, The (1995)	Action|Crime|Drama|Thriller
85	86	White Squall (1996)	Action|Adventure|Drama
88	89	Nick of Time (1995)	Action|Thriller
93	95	Broken Arrow (1996)	Action|Adventure|Thriller
96	98	Shopping (1994)	Action|Thriller
108	110	Braveheart (1995)	Action|Drama|War
110	112	Rumble in the Bronx (Hont faan kui) (1995)	Action|Adventure|Comedy|Crime

In [150…

```
movies[is_action].head(15)
```

Out[150]:

	movieId	title	genres
5	6	Heat (1995)	Action|Crime|Thriller
8	9	Sudden Death (1995)	Action
9	10	GoldenEye (1995)	Action|Adventure|Thriller
14	15	Cutthroat Island (1995)	Action|Adventure|Romance
19	20	Money Train (1995)	Action|Comedy|Crime|Drama|Thriller
22	23	Assassins (1995)	Action|Crime|Thriller
41	42	Dead Presidents (1995)	Action|Crime|Drama
43	44	Mortal Kombat (1995)	Action|Adventure|Fantasy
50	51	Guardian Angel (1994)	Action|Drama|Thriller
65	66	Lawnmower Man 2: Beyond Cyberspace (1996)	Action|Sci-Fi|Thriller
69	70	From Dusk Till Dawn (1996)	Action|Comedy|Horror|Thriller
70	71	Fair Game (1995)	Action
75	76	Screamers (1995)	Action|Sci-Fi|Thriller
77	78	Crossing Guard, The (1995)	Action|Crime|Drama|Thriller
85	86	White Squall (1996)	Action|Adventure|Drama

In [ ]: