In [3]:
```
#Movie Data set
```

In [5]:
```
import pandas as pd
```

In [7]:
```
pd.__version__
```

Out[7]: '2.2.2'

In [9]:
```
movies = pd.read_csv(r'C:\Users\mdtan\OneDrive\Desktop\NiT DS\movie.csv')
```

In [11]:
```
movies
```

Out[11]:

	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 [13]:
```
ratings = pd.read_csv(r'C:\Users\mdtan\OneDrive\Desktop\NiT DS\rating.csv')
ratings
```

Out[13]:

	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 [14]:
```
tags =  pd.read_csv(r'C:\Users\mdtan\OneDrive\Desktop\NiT DS\tag.csv')
tags
```

Out[14]:

	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 [15]:
```
print(movies.shape)
print(ratings.shape)
print(tags.shape)
```

```
(27278, 3)
(20000263, 4)
(465564, 4)
```

In [16]:
```
print(movies.columns)
print(ratings.columns)
print(tags.columns)
```

```
Index(['movieId', 'title', 'genres'], dtype='object')
Index(['userId', 'movieId', 'rating', 'timestamp'], dtype='object')
Index(['userId', 'movieId', 'tag', 'timestamp'], dtype='object')
```

In [17]:
```
del ratings['timestamp']
del tags['timestamp']
```

In [18]:
```
print(movies.columns)
print(ratings.columns)
print(tags.columns)
```

```
Index(['movieId', 'title', 'genres'], dtype='object')
Index(['userId', 'movieId', 'rating'], dtype='object')
Index(['userId', 'movieId', 'tag'], dtype='object')
```

In [19]:
```
tags.head(2)
```

Out[19]:

	userId	movieId	tag
0	18	4141	Mark Waters
1	65	208	dark hero

In [20]:
```
ratings.head(44)
```

Out[20]:

	userId	movieId	rating
0	1	2	3.5
1	1	29	3.5
2	1	32	3.5
3	1	47	3.5
4	1	50	3.5
5	1	112	3.5
6	1	151	4.0
7	1	223	4.0
8	1	253	4.0
9	1	260	4.0
10	1	293	4.0
11	1	296	4.0
12	1	318	4.0
13	1	337	3.5
14	1	367	3.5
15	1	541	4.0
16	1	589	3.5
17	1	593	3.5
18	1	653	3.0
19	1	919	3.5
20	1	924	3.5
21	1	1009	3.5
22	1	1036	4.0
23	1	1079	4.0
24	1	1080	3.5
25	1	1089	3.5
26	1	1090	4.0
27	1	1097	4.0
28	1	1136	3.5
29	1	1193	3.5
30	1	1196	4.5
31	1	1198	4.5
32	1	1200	4.0

	userId	movieId	rating
33	1	1201	3.0
34	1	1208	3.5
35	1	1214	4.0
36	1	1215	4.0
37	1	1217	3.5
38	1	1219	4.0
39	1	1222	3.5
40	1	1240	4.0
41	1	1243	3.0
42	1	1246	3.5
43	1	1249	4.0

In [21]:
```
tags.head(44)
```

Out[21]:

	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
5	65	668	bollywood
6	65	898	screwball comedy
7	65	1248	noir thriller
8	65	1391	mars
9	65	1617	neo-noir
10	65	1694	jesus
11	65	1783	noir thriller
12	65	2022	jesus
13	65	2193	dragon
14	65	2353	conspiracy theory
15	65	2662	mars
16	65	2726	noir thriller
17	65	2840	jesus
18	65	3052	jesus
19	65	5135	bollywood
20	65	6539	treasure
21	65	6874	dark hero
22	65	7013	noir thriller
23	65	7318	jesus
24	65	8529	stranded
25	65	8622	conspiracy theory
26	65	27803	Oscar (Best Foreign Language Film)
27	65	27866	New Zealand
28	65	48082	surreal
29	65	48082	unusual
30	65	51884	bollywood
31	65	58652	cute
32	65	58652	emotional

	userId	movieId	tag
33	65	58652	girls who play boys
34	65	58652	Stephen Chow
35	96	106696	animation
36	96	106696	beautiful
37	96	106696	characters
38	96	106696	Disney
39	96	106696	feminist
40	96	106696	Ice
41	96	106696	music
42	96	106696	musical
43	96	106696	pacing

In [38]:
```
movies.head(44)
```

Out[38]:

	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
12	13	Balto (1995)	Adventure|Animation|Children
13	14	Nixon (1995)	Drama
14	15	Cutthroat Island (1995)	Action|Adventure|Romance
15	16	Casino (1995)	Crime|Drama
16	17	Sense and Sensibility (1995)	Drama|Romance
17	18	Four Rooms (1995)	Comedy
18	19	Ace Ventura: When Nature Calls (1995)	Comedy
19	20	Money Train (1995)	Action|Comedy|Crime|Drama|Thriller
20	21	Get Shorty (1995)	Comedy|Crime|Thriller
21	22	Copycat (1995)	Crime|Drama|Horror|Mystery|Thriller
22	23	Assassins (1995)	Action|Crime|Thriller
23	24	Powder (1995)	Drama|Sci-Fi
24	25	Leaving Las Vegas (1995)	Drama|Romance
25	26	Othello (1995)	Drama
26	27	Now and Then (1995)	Children|Drama
27	28	Persuasion (1995)	Drama|Romance
28	29	City of Lost Children, The (Cité des enfants p...	Adventure|Drama|Fantasy|Mystery|Sci-Fi

	movieId	title	genres
29	30	Shanghai Triad (Yao a yao yao dao waipo qiao) ...	Crime|Drama
30	31	Dangerous Minds (1995)	Drama
31	32	Twelve Monkeys (a.k.a. 12 Monkeys) (1995)	Mystery|Sci-Fi|Thriller
32	33	Wings of Courage (1995)	Adventure|Romance|IMAX
33	34	Babe (1995)	Children|Drama
34	35	Carrington (1995)	Drama|Romance
35	36	Dead Man Walking (1995)	Crime|Drama
36	37	Across the Sea of Time (1995)	Documentary|IMAX
37	38	It Takes Two (1995)	Children|Comedy
38	39	Clueless (1995)	Comedy|Romance
39	40	Cry, the Beloved Country (1995)	Drama
40	41	Richard III (1995)	Drama|War
41	42	Dead Presidents (1995)	Action|Crime|Drama
42	43	Restoration (1995)	Drama
43	44	Mortal Kombat (1995)	Action|Adventure|Fantasy

In [40]:
```
# DATA STRUCTURES SERIES
```

In [42]:
```
row_0 = tags.iloc[1]
row_0
```

Out[42]:
```
userId            65
movieId          208
tag        dark hero
Name: 1, dtype: object
```

In [44]:
```
row_0 = movies.iloc[1]
row_0
```

Out[44]:
```
movieId                            2
title                 Jumanji (1995)
genres     Adventure|Children|Fantasy
Name: 1, dtype: object
```

In [46]:
```
row1=ratings.iloc[1]
row1
```

Out[46]:
```
userId      1.0
movieId    29.0
rating      3.5
Name: 1, dtype: float64
```

In [48]:
```
row2=movies.iloc[0:44]
row2
```

Out[48]:

	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
12	13	Balto (1995)	Adventure|Animation|Children
13	14	Nixon (1995)	Drama
14	15	Cutthroat Island (1995)	Action|Adventure|Romance
15	16	Casino (1995)	Crime|Drama
16	17	Sense and Sensibility (1995)	Drama|Romance
17	18	Four Rooms (1995)	Comedy
18	19	Ace Ventura: When Nature Calls (1995)	Comedy
19	20	Money Train (1995)	Action|Comedy|Crime|Drama|Thriller
20	21	Get Shorty (1995)	Comedy|Crime|Thriller
21	22	Copycat (1995)	Crime|Drama|Horror|Mystery|Thriller
22	23	Assassins (1995)	Action|Crime|Thriller
23	24	Powder (1995)	Drama|Sci-Fi
24	25	Leaving Las Vegas (1995)	Drama|Romance
25	26	Othello (1995)	Drama
26	27	Now and Then (1995)	Children|Drama
27	28	Persuasion (1995)	Drama|Romance
28	29	City of Lost Children, The (Cité des enfants p...	Adventure|Drama|Fantasy|Mystery|Sci-Fi

	movieId	title	genres
29	30	Shanghai Triad (Yao a yao yao dao waipo qiao) ...	Crime|Drama
30	31	Dangerous Minds (1995)	Drama
31	32	Twelve Monkeys (a.k.a. 12 Monkeys) (1995)	Mystery|Sci-Fi|Thriller
32	33	Wings of Courage (1995)	Adventure|Romance|IMAX
33	34	Babe (1995)	Children|Drama
34	35	Carrington (1995)	Drama|Romance
35	36	Dead Man Walking (1995)	Crime|Drama
36	37	Across the Sea of Time (1995)	Documentary|IMAX
37	38	It Takes Two (1995)	Children|Comedy
38	39	Clueless (1995)	Comedy|Romance
39	40	Cry, the Beloved Country (1995)	Drama
40	41	Richard III (1995)	Drama|War
41	42	Dead Presidents (1995)	Action|Crime|Drama
42	43	Restoration (1995)	Drama
43	44	Mortal Kombat (1995)	Action|Adventure|Fantasy

In [49]:
```
type(row_0)
```

Out[49]:
```
pandas.core.series.Series
```

In [50]:
```
type(row2)
```

Out[50]:
```
pandas.core.frame.DataFrame
```

In [51]:
```
type(row1)
```

Out[51]:
```
pandas.core.series.Series
```

In [52]:
```
print(row_0)
```

```
movieId                            2
title                 Jumanji (1995)
genres     Adventure|Children|Fantasy
Name: 1, dtype: object
```

In [53]:
```
row_0.index
```

Out[53]:
```
Index(['movieId', 'title', 'genres'], dtype='object')
```

In [56]:
```
row1['userId']
```

Out[56]:
```
1.0
```

In [60]:
```
'rating' in row_0
```

Out[60]:
```
False
```

In [62]:
```
row_0.name
```

Out[62]:
```
1
```

In [64]:
```
row_0 = row_0.rename('firstRow')
row_0.name
```

Out[64]:
```
'firstRow'
```

In [65]:
```
row_0
```

Out[65]:
```
movieId                            2
title                 Jumanji (1995)
genres     Adventure|Children|Fantasy
Name: firstRow, dtype: object
```

DATA FRAMES

In [68]:
```
tags.head()
```

Out[68]:

	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 [70]:
```
tags.index
```

Out[70]:
```
RangeIndex(start=0, stop=465564, step=1)
```

In [72]:
```
tags.columns
```

Out[72]:
```
Index(['userId', 'movieId', 'tag'], dtype='object')
```

In [74]:
```
tags.iloc[0]
```

Out[74]:
```
userId             18
movieId          4141
tag        Mark Waters
Name: 0, dtype: object
```

In [75]:
```
tags.iloc[[0,10,500]]
```

Out[75]:

	userId	movieId	tag
0	18	4141	Mark Waters
10	65	1694	jesus
500	342	55908	entirely dialogue

In [81]:
```
tags.iloc[0:44]
```

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
5	65	668	bollywood
6	65	898	screwball comedy
7	65	1248	noir thriller
8	65	1391	mars
9	65	1617	neo-noir
10	65	1694	jesus
11	65	1783	noir thriller
12	65	2022	jesus
13	65	2193	dragon
14	65	2353	conspiracy theory
15	65	2662	mars
16	65	2726	noir thriller
17	65	2840	jesus
18	65	3052	jesus
19	65	5135	bollywood
20	65	6539	treasure
21	65	6874	dark hero
22	65	7013	noir thriller
23	65	7318	jesus
24	65	8529	stranded
25	65	8622	conspiracy theory
26	65	27803	Oscar (Best Foreign Language Film)
27	65	27866	New Zealand
28	65	48082	surreal
29	65	48082	unusual
30	65	51884	bollywood
31	65	58652	cute
32	65	58652	emotional

	userId	movieId	tag
33	65	58652	girls who play boys
34	65	58652	Stephen Chow
35	96	106696	animation
36	96	106696	beautiful
37	96	106696	characters
38	96	106696	Disney
39	96	106696	feminist
40	96	106696	Ice
41	96	106696	music
42	96	106696	musical
43	96	106696	pacing

Descriptive Statistics

In [84]:
```
ratings['rating'].describe()
```

Out[84]:
```
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 [85]:
```
ratings.describe()
```

Out[85]:

	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 [86]:
```
ratings['rating'].mean()
```

Out[86]:
```
3.5255285642993797
```

In [87]:
```
ratings.mean()
```

Out[87]:
```
userId     69045.872583
movieId     9041.567330
rating         3.525529
dtype: float64
```

In [88]:
```
ratings['rating'].min()
```

Out[88]:
```
0.5
```

In [89]:
```
ratings['rating'].max()
```

Out[89]:
```
5.0
```

In [90]:
```
ratings['rating'].std()
```

Out[90]:
```
1.051988919275684
```

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

Out[91]:
```
0    4.0
Name: rating, dtype: float64
```

In [98]:
```
ratings.corr()
```

Out[98]:

	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 [101…
```
filter1 = ratings['rating'] > 10
print(filter1)
filter1.any()
```

```
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
```

Out[101…
```
False
```

In [102…
```
filter2 = ratings['rating'] > 0
filter2.all()
```

Out[102…
```
True
```

DATA CLEANING HANDLING MISSING DATA

In [104…
```
movies.shape
```

Out[104…
```
(27278, 3)
```

In [105…
```
movies.isnull().any().any()
```

Out[105…
```
False
```

In [118…
```
ratings.shape
```

Out[118…
```
(20000263, 3)
```

In [120…
```
ratings.isnull().any().any()
```

Out[120…
```
False
```

In [122…
```
tags.shape
```

Out[122…
```
(465564, 3)
```

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

Out[124…
```
True
```

In [128…
```
tags=tags.dropna()
tags
```

Out[128…

	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
…	…	…	…
465559	138446	55999	dragged
465560	138446	55999	Jason Bateman
465561	138446	55999	quirky
465562	138446	55999	sad
465563	138472	923	rise to power

465548 rows × 3 columns

In [130…
```
tags.isnull().any().any()
```

Out[130…
```
False
```

In [132…
```
tags.shape
```

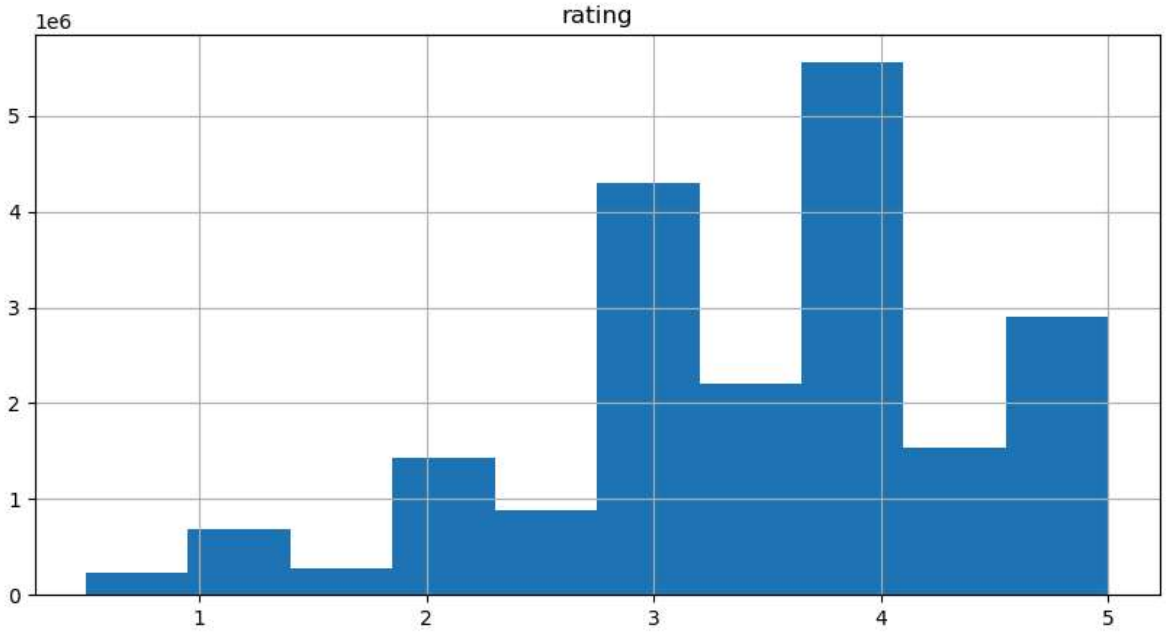
Out[132…
```
(465548, 3)
```

DATA VISUALIZATION

In [135…
```
%matplotlib inline
```

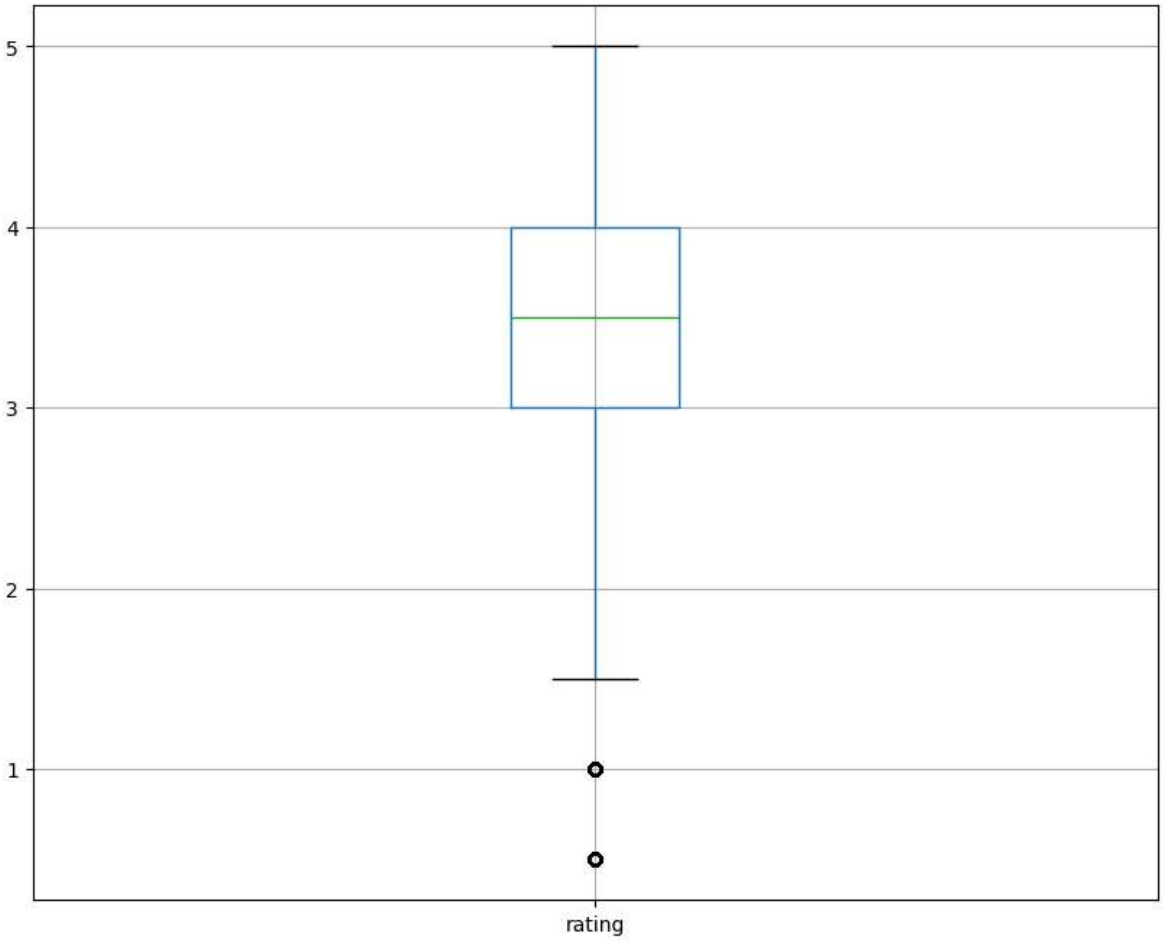
In [158…
```
ratings.hist(column='rating',figsize=(10,5))
```

Out[158…
```
array([[<Axes: title={'center': 'rating'}>]], dtype=object)
```



In [143…
```
ratings.boxplot(column='rating',figsize=(10,8))
```

Out[143…
```
<Axes: >
```



Slicing Out Columns

In [148…
```
tags['tag'].head()
```

Out[148…
```
0      Mark Waters
1        dark hero
2        dark hero
3    noir thriller
4        dark hero
Name: tag, dtype: object
```

In [150…
```
movies[['title','genres']].head()
```

Out[150…

	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 [152…
```
ratings[-10:]
```

Out[152…

	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 [164…

```
tag_counts=tags['tag'].value_counts()
tag_counts[-10:]
```

Out[164…

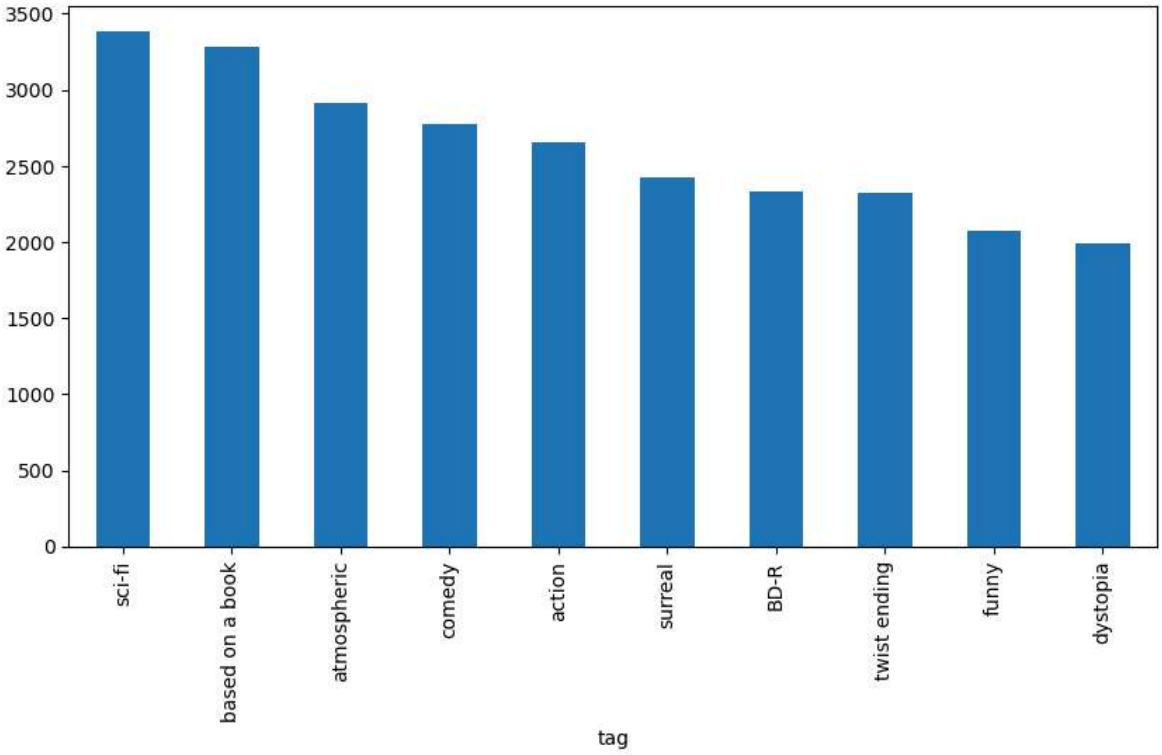
```
tag
missing child                   1
Ron Moore                       1
Citizen Kane                    1
mullet                          1
biker gang                      1
Paul Adelstein                  1
the wig                         1
killer fish                     1
genetically modified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

In [170…

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

Out[170…

```
<Axes: xlabel='tag'>
```


3500
3000
2500
2000
1500
1000
500
0
sci-fi
based on a book
atmospheric
comedy
action
surreal
BD-R
twist ending
funny
dystopia
tag

In [ ]: