

```
In [2]: import pandas as pd
```

```
In [8]: movies=pd.read_csv(r"C:\Users\S SHYAMILI\OneDrive\Desktop\data science\my projects\movies.head(20)
```

Out[8]:

	movielid	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

```
In [33]: ratings=pd.read_csv(r"C:\Users\S SHYAMILI\OneDrive\Desktop\data science\my projects\ratings.head()
```

```
Out[33]:
```

	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

```
In [149...:
```

```
tags=pd.read_csv(r"C:\Users\S SHYAMILI\OneDrive\Desktop\data science\my projects\rat
tags.head(5)
```

```
Out[149...:
```

	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

```
In [35]:
```

```
del ratings['timestamp']
del tags['timestamp']
```

```
In [41]:
```

```
row=tags.iloc[0]
type(row)
```

```
Out[41]:
```

pandas.core.series.Series

```
In [43]:
```

```
row
```

```
Out[43]:
```

```
userId      1.0
movieId     2.0
rating      3.5
Name: 0, dtype: float64
```

```
In [45]:
```

```
row.index
```

```
Out[45]:
```

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

```
In [47]:
```

```
row.index
```

```
Out[47]:
```

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

```
In [49]:
```

```
row['userId']
```

```
Out[49]:
```

1.0

```
In [51]: 'rating' in row
```

```
Out[51]: True
```

```
In [53]: row.name
```

```
Out[53]: 0
```

```
In [57]: row=row.rename('first row')
row.name
```

```
Out[57]: 'first row'
```

```
In [59]: tags.head()
```

```
Out[59]:   userId  movieId  rating
```

	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

```
In [61]: tags.index
```

```
Out[61]: RangeIndex(start=0, stop=20000263, step=1)
```

```
In [65]: tags.columns
```

```
Out[65]: Index(['userId', 'movieId', 'rating'], dtype='object')
```

```
In [69]: tags.iloc[[10,35,55,60]]
```

```
Out[69]:   userId  movieId  rating
```

	userId	movieId	rating
10	1	293	4.0
35	1	1214	4.0
55	1	1350	3.5
60	1	1525	3.0

```
In [71]: ratings['rating'].describe()
```

```
Out[71]: 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 [77]: ratings.describe()
```

```
Out[77]:      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 [79]: ratings['rating'].mean()
```

```
Out[79]: 3.5255285642993797
```

```
In [81]: ratings['rating'].min()
```

```
Out[81]: 0.5
```

```
In [83]: ratings['rating'].max()
```

```
Out[83]: 5.0
```

```
In [85]: ratings['rating'].mode()
```

```
Out[85]: 0    4.0
          Name: rating, dtype: float64
```

```
In [87]: ratings['rating'].std()
```

```
Out[87]: 1.051988919275684
```

```
In [91]: ratings.corr()
```

```
Out[91]:
```

	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 [95]: filter=ratings['rating']>10  
print(filter)  
filter.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[95]: <bound method Series.any of 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 [105...]: filter1=ratings['rating']>0  
filter.all()
```

```
Out[105...]: False
```

```
In [107...]: movies.shape
```

```
Out[107...]: (27278, 3)
```

```
In [109...]: movies.isnull().any().all()
```

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

```
In [113...]: ratings.shape
```

```
Out[113...]: (20000263, 3)
```

```
In [117... ratings.isnull().any().all()
```

```
Out[117... False
```

```
In [119... tags.shape
```

```
Out[119... (20000263, 3)
```

```
In [121... tags.isnull().any().all()
```

```
Out[121... False
```

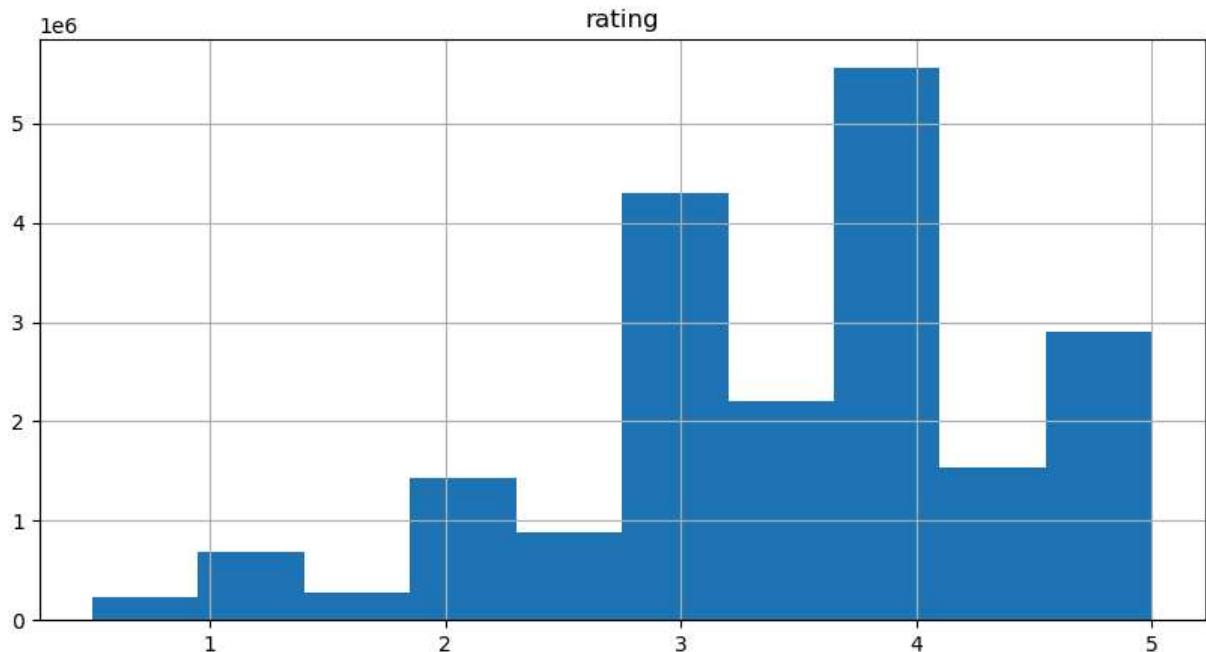
```
In [125... #if as null values use .dropna()
```

```
In [127... import matplotlib.pyplot as plt
```

```
In [129... %matplotlib inline
```

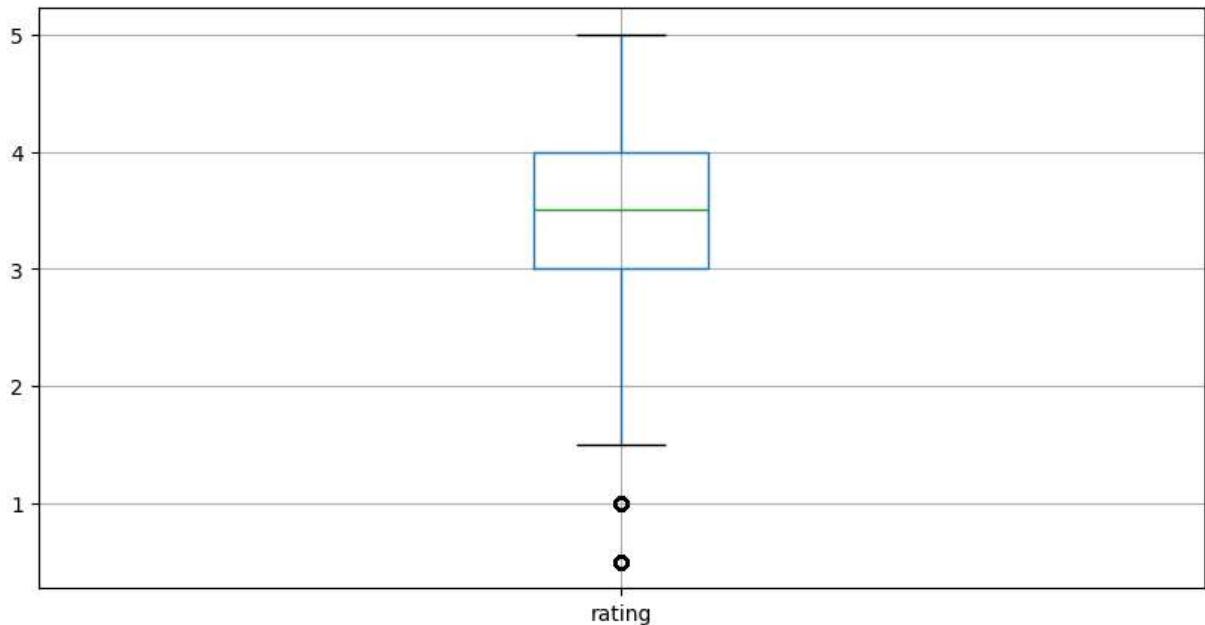
```
In [145... ratings.hist(column='rating', figsize=(10,5))
```

```
Out[145... array([[[<Axes: title={'center': 'rating'}>]], dtype=object)]
```



```
In [141... ratings.boxplot(column='rating', figsize=(10,5))
```

```
Out[141... <Axes: >
```



```
In [156...]: tags[['userId']].head()
```

```
Out[156...]:
```

0	1
1	1
2	1
3	1
4	1

Name: userId, dtype: int64

```
In [166...]: movies[['title', 'genres']].head()
```

```
Out[166...]:
```

	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 [168...]: ratings[-10:]
```

Out[168...]

	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 [182...]

```
tag_count=tags['userId'].value_counts()  
tag_count[:-10]
```

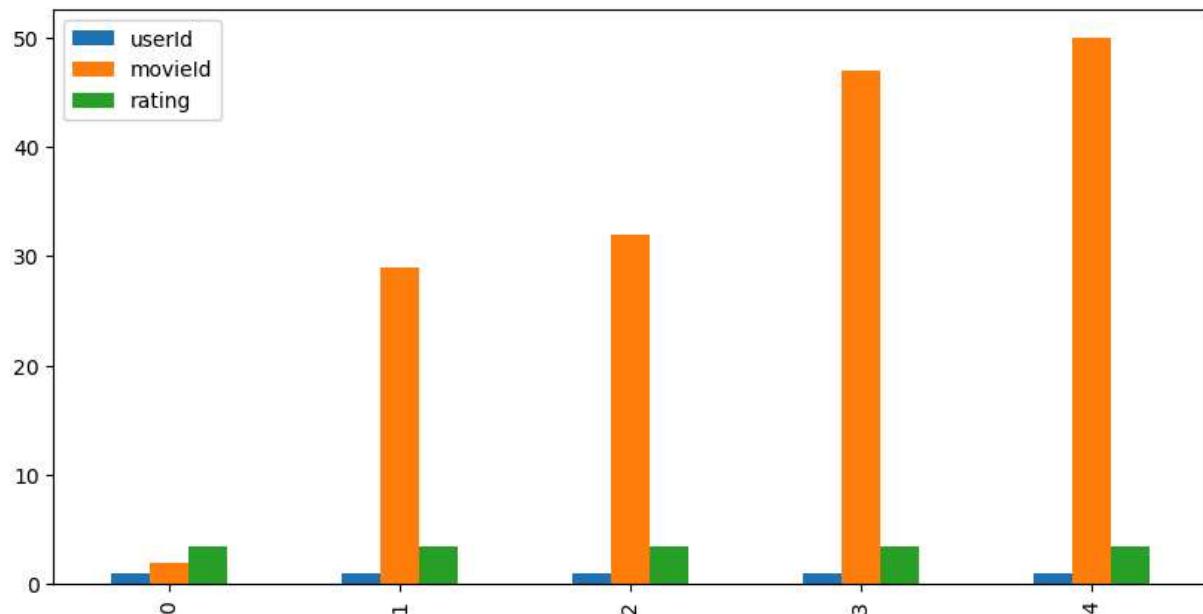
Out[182...]

```
userId  
118205    9254  
8405      7515  
82418     5646  
121535    5520  
125794    5491  
...  
79817      20  
75748      20  
53591      20  
5205       20  
43401      20  
Name: count, Length: 138483, dtype: int64
```

In [184...]

```
tags.head().plot(kind='bar', figsize=(10,5))
```

Out[184...]



```
In [192]:  
is_highly_rated=ratings['rating']>=5.0  
ratings[is_highly_rated][30:50]
```

Out[192...]

	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

In [196...]

```
is_action=movies['genres'].str.contains("Action")
movies[is_action][300:400]
```

Out[196...]

	movield	title	genres
2249	2334	Siege, The (1998)	Action Thriller
2259	2344	Runaway Train (1985)	Action Adventure Drama Thriller
2268	2353	Enemy of the State (1998)	Action Thriller
2279	2364	Godzilla 1985: The Legend Is Reborn (Gojira) (...)	Action Horror Sci-Fi Thriller
2280	2365	King Kong vs. Godzilla (Kingukongu tai Gojira)...	Action Sci-Fi
...
2858	2944	Dirty Dozen, The (1967)	Action Drama War
2861	2947	Goldfinger (1964)	Action Adventure Thriller
2862	2948	From Russia with Love (1963)	Action Adventure Thriller
2863	2949	Dr. No (1962)	Action Adventure Thriller
2865	2951	Fistful of Dollars, A (Per un pugno di dollari...)	Action Western

100 rows × 3 columns

In [198...]

`movies[is_action].head()`

Out[198...]

	movield	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

In [206...]

`ratings_count = ratings[['movieId','rating']].groupby('movieId').count()
ratings_count`

Out[206...]

rating**movield**

1	49695
2	22243
3	12735
4	2756
5	12161
...	...
131254	1
131256	1
131258	1
131260	1
131262	1

26744 rows × 1 columns

In [208...]

```
average_rating = ratings[['movieId','rating']].groupby('movieId').mean()  
average_rating.head()
```

Out[208...]

rating**movield**

1	3.921240
2	3.211977
3	3.151040
4	2.861393
5	3.064592

In [210...]

```
movie_count = ratings[['movieId','rating']].groupby('movieId').count()  
movie_count.head()
```

Out[210...]

rating**movield**

1	49695
2	22243
3	12735
4	2756
5	12161

In [212...]

```
movie_count = ratings[['movieId','rating']].groupby('movieId').count()  
movie_count.tail()
```

Out[212...]

rating**movield**

131254	1
131256	1
131258	1
131260	1
131262	1

In [214...]

```
tags.head()
```

Out[214...]

userId movield 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

In [216...]

```
movies.head()
```

Out[216...]

	moviedb	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

In [219...]

```
t = movies.merge(tags, on='movieId', how='inner')
t.head()
```

Out[219...]

	moviedb	title	genres	userId	rating	timestamp
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	3	4.0	1999-12-11 13:36:47
1	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	6	5.0	1997-03-13 17:50:52
2	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	8	4.0	1996-06-05 13:37:51
3	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	10	4.0	1999-11-25 02:44:47
4	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	11	4.5	2009-01-02 01:13:41

In [223...]

```
avg_ratings= ratings.groupby('movieId', as_index=False).mean()
del avg_ratings['userId']
avg_ratings.head()
```

Out[223...]

	moviedb	rating
0	1	3.921240
1	2	3.211977
2	3	3.151040
3	4	2.861393
4	5	3.064592

In [225...]

```
box_office = movies.merge(avg_ratings, on='movieId', how='inner')
box_office.tail()
```

Out[225...]

	movield	title	genres	rating
26739	131254	Kein Bund für's Leben (2007)	Comedy	4.0
26740	131256	Feuer, Eis & Dosenbier (2002)	Comedy	4.0
26741	131258	The Pirates (2014)	Adventure	2.5
26742	131260	Rentun Ruusu (2001)	(no genres listed)	3.0
26743	131262	Innocence (2014)	Adventure Fantasy Horror	4.0

In [227...]

```
is_highly Rated = box_office['rating'] >= 4.0
box_office[is_highly Rated][-5:]
```

Out[227...]

	movield	title	genres	rating
26737	131250	No More School (2000)	Comedy	4.0
26738	131252	Forklift Driver Klaus: The First Day on the Jo...	Comedy Horror	4.0
26739	131254	Kein Bund für's Leben (2007)	Comedy	4.0
26740	131256	Feuer, Eis & Dosenbier (2002)	Comedy	4.0
26743	131262	Innocence (2014)	Adventure Fantasy Horror	4.0

In [229...]

```
is_Adventure = box_office['genres'].str.contains('Adventure')
box_office[is_Adventure][:5]
```

Out[229...]

	movield	title	genres	rating
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	3.921240
1	2	Jumanji (1995)	Adventure Children Fantasy	3.211977
7	8	Tom and Huck (1995)	Adventure Children	3.142049
9	10	GoldenEye (1995)	Action Adventure Thriller	3.430029
12	13	Balto (1995)	Adventure Animation Children	3.272416

In [231...]

```
box_office[is_Adventure & is_highly Rated][-5:]
```

Out[231...]

	movield	title	genres	rating
26611	130586	Itinerary of a Spoiled Child (1988)	Adventure Drama	4.5
26655	130996	The Beautiful Story (1992)	Adventure Drama Fantasy	5.0
26667	131050	Stargate SG-1 Children of the Gods - Final Cut...	Adventure Sci-Fi Thriller	5.0
26736	131248	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	4.0
26743	131262	Innocence (2014)	Adventure Fantasy Horror	4.0

In [233...]

```
movie_genres = movies['genres'].str.split('|', expand=True)
```

In [235...]

```
movie_genres[:10]
```

Out[235...]

	0	1	2	3	4	5	6	7	8	9
0	Adventure	Animation	Children	Comedy	Fantasy	None	None	None	None	None
1	Adventure	Children	Fantasy	None	None	None	None	None	None	None
2	Comedy	Romance	None	None	None	None	None	None	None	None
3	Comedy	Drama	Romance	None	None	None	None	None	None	None
4	Comedy	None	None	None	None	None	None	None	None	None
5	Action	Crime	Thriller	None	None	None	None	None	None	None
6	Comedy	Romance	None	None	None	None	None	None	None	None
7	Adventure	Children	None	None	None	None	None	None	None	None
8	Action	None	None	None	None	None	None	None	None	None
9	Action	Adventure	Thriller	None	None	None	None	None	None	None

In [237...]

```
movie_genres['isComedy'] = movies['genres'].str.contains('Comedy')
```

In [239...]

```
movie_genres[:10]
```

Out[239...]

	0	1	2	3	4	5	6	7	8	9	isCc
0	Adventure	Animation	Children	Comedy	Fantasy	None	None	None	None	None	None
1	Adventure	Children	Fantasy	None	None	None	None	None	None	None	None
2	Comedy	Romance	None	None	None	None	None	None	None	None	None
3	Comedy	Drama	Romance	None	None	None	None	None	None	None	None
4	Comedy	None	None	None	None	None	None	None	None	None	None
5	Action	Crime	Thriller	None	None	None	None	None	None	None	None
6	Comedy	Romance	None	None	None	None	None	None	None	None	None
7	Adventure	Children	None	None	None	None	None	None	None	None	None
8	Action	None	None	None	None	None	None	None	None	None	None
9	Action	Adventure	Thriller	None	None	None	None	None	None	None	None



In []: