In [1]:

```
import pandas as pd
```

In [2]:

```
tags = pd.read_csv(r"C:\Users\Vansh\Downloads\archive\archive\tag.csv")
tags
```

Out[2]:

	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 [4]:

```
tags.shape
```

Out[4]:

```
(465564, 4)
```

In [7]:

```
tags.head()
```

Out[7]:

	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

In [9]:

```
movies = pd.read_csv(r"C:\Users\Vansh\Downloads\archive\archive\movie.csv")
movies
```

Out[9]:

	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 [11]:

```
movies.shape
```

Out[11]:

```
(27278, 3)
```

In [13]:

```
movies.head()
```

Out[13]:

	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

In [15]:

```
movies.tail()
```

Out[15]:

	movieId	title	genres
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

In [17]:

```
movies[:]
```

Out[17]:

	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 [19]:

```
tags = pd.read_csv(r"C:\Users\Vansh\Downloads\archive\archive\tag.csv")
tags
```

Out[19]:

	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 [21]:

```
ratings = pd.read_csv(r"C:\Users\Vansh\Downloads\archive\archive\movie.csv")
dates = (['TIMESTAMP'])
ratings.head()
```

Out[21]:

	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

In [23]:

```
ratings.shape
```

Out[23]:

```
(27278, 3)
```

In [25]:

```
movies = pd.read_csv(r"C:\Users\Vansh\Downloads\archive\archive\movie.csv")
print(type(movies))
movies.head(20)
```

```
<class 'pandas.core.frame.DataFrame'>
```

Out[25]:

	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

In [27]:

```
ratings.columns
```

Out[27]:

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

In [29]:

```
tags.columns
```

Out[29]:

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

In [31]:

```
ratings = pd.read_csv(r'C:\Users\Vansh\Downloads\archive\archive\movie.csv')#,pa
ratings.head()
```

Out[31]:

	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

DATA STRUCTURES

In [34]:

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

Out[34]:

```
pandas.core.series.Series
```

In [36]:

```
print(row_0)
```

```
userId                        18
movieId                     4141
tag                  Mark Waters
timestamp    2009-04-24 18:19:40
Name: 0, dtype: object
```

In [38]:

```
row_0.index
```

Out[38]:

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

In [40]:

```
row_0['userId']
```

Out[40]:

```
18
```

In [42]:

```
'rating' in row_0
```

Out[42]:

```
False
```

In [44]:

```
row_0.name
```

Out[44]:

```
0
```

In [46]:

```
row_0 = row_0.rename('firstRow')
row_0.name
```

Out[46]:

```
'firstRow'
```

In [48]:

```
row_0 = row_0.rename('firstRow')
row_0.name
```

Out[48]:

```
'firstRow'
```

DATA FRAMES

```
In [51]: tags.head()
```

Out[51]:

	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

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

```
Out[53]: RangeIndex(start=0, stop=465564, step=1)
```

```
In [55]: tags.columns
```

```
Out[55]: Index(['userId', 'movieId', 'tag', 'timestamp'], dtype='object')
```

```
In [57]: tags.iloc[[ 0,11,500]]
```

Out[57]:

	userId	movieId	tag	timestamp
0	18	4141	Mark Waters	2009-04-24 18:19:40
11	65	1783	noir thriller	2013-05-10 01:39:43
500	342	55908	entirely dialogue	2012-01-31 18:41:16

DESCRIPTIVE STASTICS

```
In [60]: ratings
```

Out[60]:

	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 [62]: ratings['movieId'].describe()
```

```
Out[62]: count     27278.000000
         mean      59855.480570
         std       44429.314697
         min           1.000000
         25%        6931.250000
         50%       68068.000000
         75%      100293.250000
         max      131262.000000
         Name: movieId, dtype: float64
```

```
In [64]: ratings.columns
```

```
Out[64]: Index(['movieId', 'title', 'genres'], dtype='object')
```

```
In [66]: ratings.describe()
```

Out[66]:

	movieId
count	27278.000000
mean	59855.480570
std	44429.314697
min	1.000000
25%	6931.250000
50%	68068.000000
75%	100293.250000
max	131262.000000

In [68]:

```python
ratings.head()
```

Out[68]:

	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

In [72]:

```python
ratings.columns
```

Out[72]:

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

In [76]:

```python
ratings.describe()
```

Out[76]:

	movieId
count	27278.000000
mean	59855.480570
std	44429.314697
min	1.000000
25%	6931.250000
50%	68068.000000
75%	100293.250000
max	131262.000000

In [78]:

```python
ratings.head()
```

Out[78]:

	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

In [80]:
```
ratings['movieId'].mean()
```

Out[80]:
```
59855.48057042305
```

In [82]:
```
ratings['movieId'].max()
```

Out[82]:
```
131262
```

In [84]:
```
ratings['movieId'].mode()
```

Out[84]:
```
0             1
1             2
2             3
3             4
4             5
          ...  
27273    131254
27274    131256
27275    131258
27276    131260
27277    131262
Name: movieId, Length: 27278, dtype: int64
```

In [86]:
```
filter1 = ratings['movieId'] > 10
print(filter1)
filter1.any()
```

```
0        False
1        False
2        False
3        False
4        False
         ...  
27273     True
27274     True
27275     True
27276     True
27277     True
Name: movieId, Length: 27278, dtype: bool
```

Out[86]:
```
True
```

In [88]:
```
filter2 = ratings['movieId'] > 0
filter2.all()
```

Out[88]:
```
True
```

DATA CLEANING : HANDLING MISSING DATA

```
In [91]:  movies.shape
```

```
Out[91]:  (27278, 3)
```

```
In [93]:  movies.isnull().any().any()
```

```
Out[93]:  False
```

```
In [97]:  ratings.shape
```

```
Out[97]:  (27278, 3)
```

```
In [99]:  tags.isnull().any().any()
```

```
Out[99]:  True
```

```
In [101…  tags = tags.dropna()
          tags
```

Out[101…

	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

465548 rows × 4 columns

```
In [103…  tags.isnull().any().any()
```

```
Out[103…  False
```

```
In [105…  tags.shape
```

```
Out[105…  (465548, 4)
```

DATA VISUALIZATION

In [109…

```
import matplotlib.pyplot as plt
```

In [111…

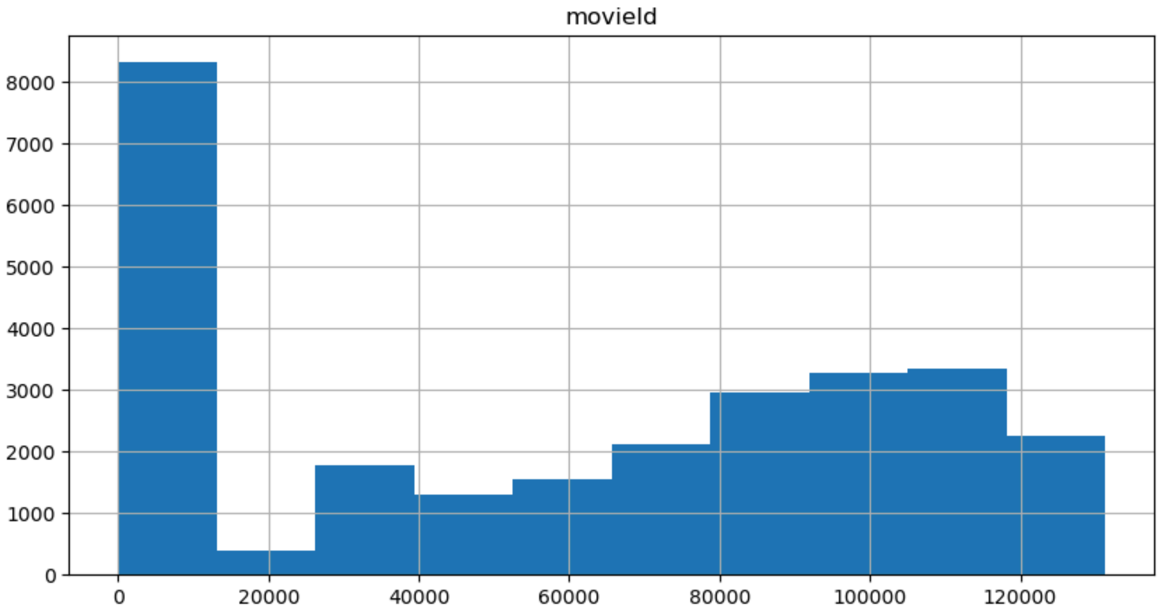
```
 %matplotlib inline

ratings.hist(column='movieId', figsize=(10,5))
```

Out[111…

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

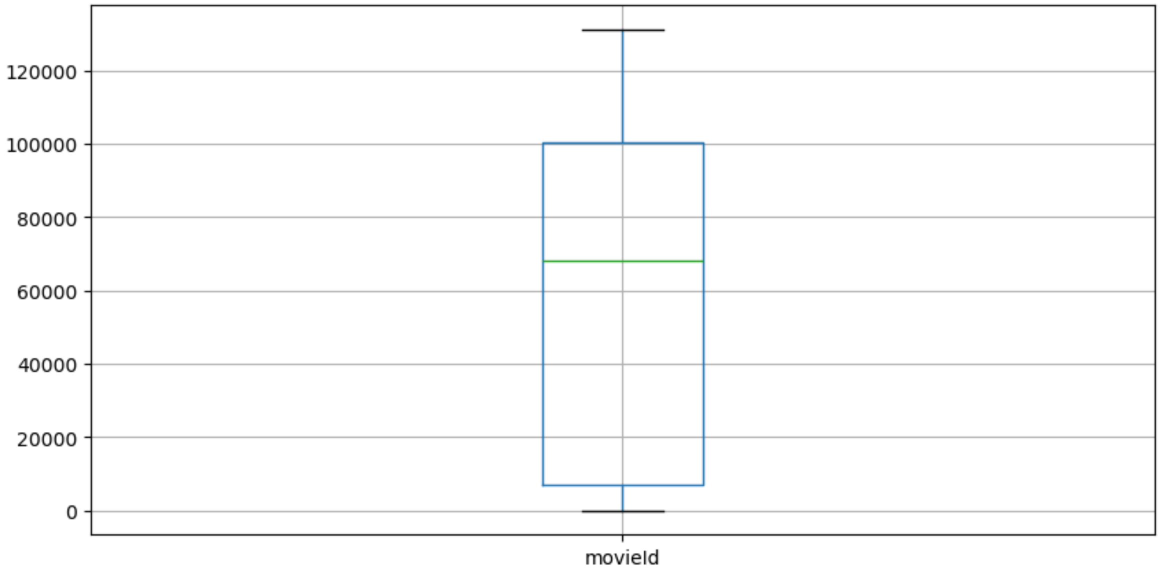
In [113…

```
plt.show()
```



In [115…

```
ratings.boxplot(column='movieId',figsize=(10,5))
plt.show()
```



SLICING OUT COLUMNS

In [118…

```
tags['tag'].head()
```

Out[118…

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

In [120…

```
movies[['movieId','genres']].head()
```

Out[120…

	movieId	genres
0	1	Adventure|Animation|Children|Comedy|Fantasy
1	2	Adventure|Children|Fantasy
2	3	Comedy|Romance
3	4	Comedy|Drama|Romance
4	5	Comedy

In [122…

```
ratings[-10:]
```

Out[122…

	movieId	title	genres
27268	131241	Ants in the Pants (2000)	Comedy|Romance
27269	131243	Werner - Gekotzt wird später (2003)	Animation|Comedy
27270	131248	Brother Bear 2 (2006)	Adventure|Animation|Children|Comedy|Fantasy
27271	131250	No More School (2000)	Comedy
27272	131252	Forklift Driver Klaus: The First Day on the Jo...	Comedy|Horror
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

In [128…

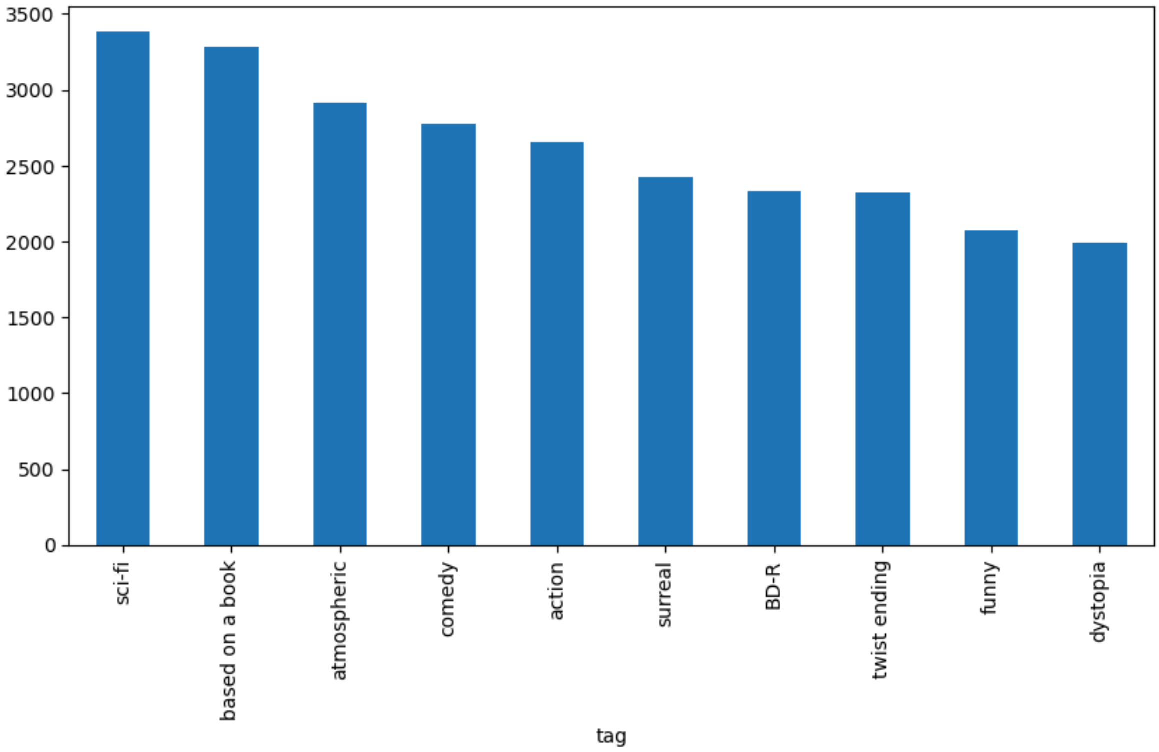
```
tag_counts = tags['tag'].value_counts()
tag_counts
```

Out[128…

```
tag
sci-fi                          3384
based on a book                 3281
atmospheric                     2917
comedy                          2779
action                          2657
                                ... 
Paul Adelstein                     1
the wig                            1
killer fish                        1
genetically modified monsters      1
topless scene                      1
Name: count, Length: 38643, dtype: int64
```

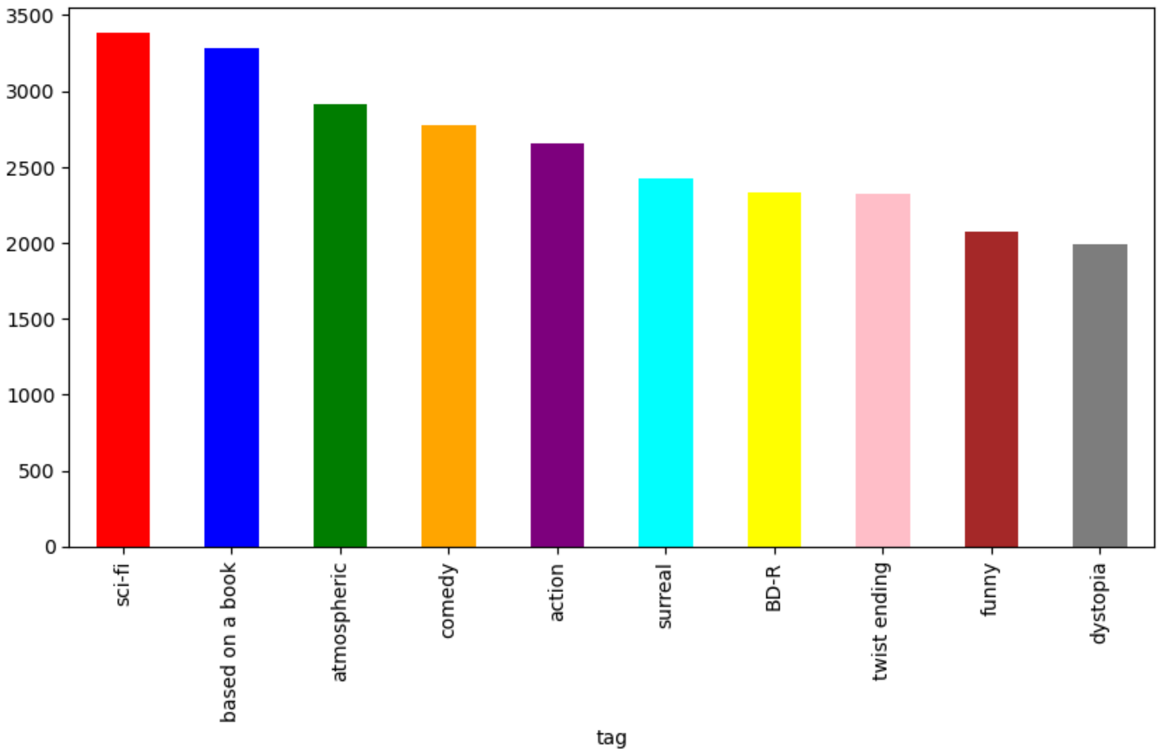
In [135…

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



In [132…

```
tag_counts[:10].plot(kind='bar', figsize=(10, 5), color=['red', 'blue', 'green',
plt.show()
```



FILTER FOR SELECTING ROWS

In [137…

```
is_highly_rated = ratings['movieId'] >= 5.0
ratings[is_highly_rated][30:50]
```

Out[137…

	movieId	title	genres
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
44	45	To Die For (1995)	Comedy|Drama|Thriller
45	46	How to Make an American Quilt (1995)	Drama|Romance
46	47	Seven (a.k.a. Se7en) (1995)	Mystery|Thriller
47	48	Pocahontas (1995)	Animation|Children|Drama|Musical|Romance
48	49	When Night Is Falling (1995)	Drama|Romance
49	50	Usual Suspects, The (1995)	Crime|Mystery|Thriller
50	51	Guardian Angel (1994)	Action|Drama|Thriller
51	52	Mighty Aphrodite (1995)	Comedy|Drama|Romance
52	53	Lamerica (1994)	Adventure|Drama
53	54	Big Green, The (1995)	Children|Comedy

In [139…

```
is_action = movies['genres'].str.contains('Action')
movies[is_action][5:15]
```

Out[139…

	movieId	title	genres
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 [143…

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

Out[143…

	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

GROUP BY AND AGGREGATE

In [148…

```
print(ratings.columns)
```

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

In [209…

```
ratings_count = ratings.groupby("movieId").count()
ratings_count
```

Out[209…

	title	genres
movieId		
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
...	...	...
131254	1	1
131256	1	1
131258	1	1
131260	1	1
131262	1	1

27278 rows × 2 columns

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: