```python
In [1]: import pandas as pd

In [76]: df=pd.read_csv(r"C:\Users\ymani\Dropbox\PC\Downloads\merged_dataset.csv")

In [78]: df
```

Out[78]:

	name	year	movie_rated	run_length	genres	release_date	rating	num_
0	Inception	2010	PG-13	2h 28min	Action; Adventure; Sci-Fi;	16 July 2010 (USA)	8.8	19
1	The Lord of the Rings: The Fellowship of the Ring	2001	PG-13	2h 58min	Action; Adventure; Drama;	19 December 2001 (USA)	8.8	16
2	The Lord of the Rings: The Return of the King	2003	PG-13	3h 21min	Adventure; Drama; Fantasy;	17 December 2003 (USA)	8.9	15
3	The Dark Knight Rises	2012	PG-13	2h 44min	Action; Adventure;	20 July 2012 (USA)	8.4	14
4	The Lord of the Rings: The Two Towers	2002	PG-13	2h 59min	Adventure; Drama; Fantasy;	18 December 2002 (USA)	8.7	14
...	...	...	...	...	...	...	...	
1495	Catch-22	1970	R	2h 2min	Comedy; Drama; War;	24 June 1970 (USA)	7.1	
1496	The Great Raid	2005	R	2h 12min	Action; Drama; War;	12 August 2005 (USA)	6.7	
1497	Saints and Soldiers	2003	PG-13	1h 30min	Action; Drama; War;	25 March 2005 (USA)	6.7	
1498	Stop-Loss	2008	R	1h 52min	Drama; War;	28 March 2008 (USA)	6.4	
1499	The Alamo	2004	PG-13	2h 17min	Drama; History; War;	9 April 2004 (USA)	6.0	

1500 rows × 9 columns

```
In [80]: df.head()
```

Out[80]:

	name	year	movie_rated	run_length	genres	release_date	rating	num_rate
0	Inception	2010	PG-13	2h 28min	Action; Adventure; Sci-Fi;	16 July 2010 (USA)	8.8	19816
1	The Lord of the Rings: The Fellowship of the Ring	2001	PG-13	2h 58min	Action; Adventure; Drama;	19 December 2001 (USA)	8.8	16091
2	The Lord of the Rings: The Return of the King	2003	PG-13	3h 21min	Adventure; Drama; Fantasy;	17 December 2003 (USA)	8.9	15938
3	The Dark Knight Rises	2012	PG-13	2h 44min	Action; Adventure;	20 July 2012 (USA)	8.4	14703
4	The Lord of the Rings: The Two Towers	2002	PG-13	2h 59min	Adventure; Drama; Fantasy;	18 December 2002 (USA)	8.7	14401

```
In [82]: df.head(3)
```

Out[82]:

	name	year	movie_rated	run_length	genres	release_date	rating	num_rate
0	Inception	2010	PG-13	2h 28min	Action; Adventure; Sci-Fi;	16 July 2010 (USA)	8.8	19816
1	The Lord of the Rings: The Fellowship of the Ring	2001	PG-13	2h 58min	Action; Adventure; Drama;	19 December 2001 (USA)	8.8	16091
2	The Lord of the Rings: The Return of the King	2003	PG-13	3h 21min	Adventure; Drama; Fantasy;	17 December 2003 (USA)	8.9	15938

```
In [84]: df.tail()
```

Out[84]:

	name	year	movie_rated	run_length	genres	release_date	rating	num_rate
1495	Catch-22	1970	R	2h 2min	Comedy; Drama; War;	24 June 1970 (USA)	7.1	2142
1496	The Great Raid	2005	R	2h 12min	Action; Drama; War;	12 August 2005 (USA)	6.7	2096
1497	Saints and Soldiers	2003	PG-13	1h 30min	Action; Drama; War;	25 March 2005 (USA)	6.7	1973
1498	Stop-Loss	2008	R	1h 52min	Drama; War;	28 March 2008 (USA)	6.4	1945
1499	The Alamo	2004	PG-13	2h 17min	Drama; History; War;	9 April 2004 (USA)	6.0	1929

```
In [86]: df=df.rename(columns={'run_length':'length'})
```

```
In [88]: df
```

Out[88]:

	name	year	movie_rated	length	genres	release_date	rating	num_rater
0	Inception	2010	PG-13	2h 28min	Action; Adventure; Sci-Fi;	16 July 2010 (USA)	8.8	198167
1	The Lord of the Rings: The Fellowship of the Ring	2001	PG-13	2h 58min	Action; Adventure; Drama;	19 December 2001 (USA)	8.8	160916
2	The Lord of the Rings: The Return of the King	2003	PG-13	3h 21min	Adventure; Drama; Fantasy;	17 December 2003 (USA)	8.9	159385
3	The Dark Knight Rises	2012	PG-13	2h 44min	Action; Adventure;	20 July 2012 (USA)	8.4	147032
4	The Lord of the Rings: The Two Towers	2002	PG-13	2h 59min	Adventure; Drama; Fantasy;	18 December 2002 (USA)	8.7	144018
...	...	...	...	...	...	...	...	
1495	Catch-22	1970	R	2h 2min	Comedy; Drama; War;	24 June 1970 (USA)	7.1	2142
1496	The Great Raid	2005	R	2h 12min	Action; Drama; War;	12 August 2005 (USA)	6.7	2096
1497	Saints and Soldiers	2003	PG-13	1h 30min	Action; Drama; War;	25 March 2005 (USA)	6.7	1973
1498	Stop-Loss	2008	R	1h 52min	Drama; War;	28 March 2008 (USA)	6.4	1945
1499	The Alamo	2004	PG-13	2h 17min	Drama; History; War;	9 April 2004 (USA)	6.0	1929

1500 rows × 9 columns

In [94]: `df=df.rename(columns={'name':'Name'})`

In [96]: `df['length']`

```
Out[96]: 0        2h 28min
         1        2h 58min
         2        3h 21min
         3        2h 44min
         4        2h 59min
                    ...
         1495      2h 2min
         1496     2h 12min
         1497     1h 30min
         1498     1h 52min
         1499     2h 17min
         Name: length, Length: 1500, dtype: object
```

```
In [98]: df['Name']
```

```
Out[98]: 0                                         Inception
         1        The Lord of the Rings: The Fellowship of the Ring
         2          The Lord of the Rings: The Return of the King
         3                                  The Dark Knight Rises
         4             The Lord of the Rings: The Two Towers
                                      ...
         1495                                       Catch-22
         1496                                  The Great Raid
         1497                              Saints and Soldiers
         1498                                      Stop-Loss
         1499                                      The Alamo
         Name: Name, Length: 1500, dtype: object
```

```
In [104… fiv=df.head()
```

```
In [108… fiv
```

	Name	year	movie_rated	length	genres	release_date	rating	num_raters
0	Inception	2010	PG-13	2h 28min	Action; Adventure; Sci-Fi;	16 July 2010 (USA)	8.8	1981675
1	The Lord of the Rings: The Fellowship of the Ring	2001	PG-13	2h 58min	Action; Adventure; Drama;	19 December 2001 (USA)	8.8	1609165
2	The Lord of the Rings: The Return of the King	2003	PG-13	3h 21min	Adventure; Drama; Fantasy;	17 December 2003 (USA)	8.9	1593859
3	The Dark Knight Rises	2012	PG-13	2h 44min	Action; Adventure;	20 July 2012 (USA)	8.4	1470329
4	The Lord of the Rings: The Two Towers	2002	PG-13	2h 59min	Adventure; Drama; Fantasy;	18 December 2002 (USA)	8.7	1440188

Descriptive Statistics

```
fiv['rating'].mean()
```

Out[110…   8.719999999999999

```
df['rating'].median()
```

Out[112…   7.6

```
fiv['rating'].median()
```

Out[114…   8.8

```
fiv[['rating','num_raters']].median()
```

Out[116…
```
rating              8.8
num_raters    1593859.0
dtype: float64
```

```
df[['num_raters','num_reviews']].describe()
```

	num_raters	num_reviews
count	1.500000e+03	1500.000000
mean	4.116746e+05	996.054667
std	3.181686e+05	1009.968589
min	1.929000e+04	102.000000
25%	1.912445e+05	424.750000
50%	3.414520e+05	724.500000
75%	5.654030e+05	1223.250000
max	2.258845e+06	10279.000000

```python
fiv[['rating','num_reviews']].describe()
```

	rating	num_reviews
count	5.000000	5.000000
mean	8.720000	3680.800000
std	0.192354	1073.180414
min	8.400000	2559.000000
25%	8.700000	2979.000000
50%	8.800000	3681.000000
75%	8.800000	3820.000000
max	8.900000	5365.000000

```python
fiv['num_reviews'].skew()
```

1.0189511787076169

```python
fiv.agg({'rating':['min','max','mean','skew','count'],'num_reviews':['min','skew
```

	rating	num_reviews
min	8.400000	2559.000000
max	8.900000	5365.000000
mean	8.720000	3680.800000
skew	-1.517474	1.018951
count	5.000000	NaN
median	NaN	3681.000000

```python
df.head()
```

	Name	year	movie_rated	length	genres	release_date	rating	num_raters
0	Inception	2010	PG-13	2h 28min	Action; Adventure; Sci-Fi;	16 July 2010 (USA)	8.8	1981675
1	The Lord of the Rings: The Fellowship of the Ring	2001	PG-13	2h 58min	Action; Adventure; Drama;	19 December 2001 (USA)	8.8	1609165
2	The Lord of the Rings: The Return of the King	2003	PG-13	3h 21min	Adventure; Drama; Fantasy;	17 December 2003 (USA)	8.9	1593859
3	The Dark Knight Rises	2012	PG-13	2h 44min	Action; Adventure;	20 July 2012 (USA)	8.4	1470329
4	The Lord of the Rings: The Two Towers	2002	PG-13	2h 59min	Adventure; Drama; Fantasy;	18 December 2002 (USA)	8.7	1440188

```python
fiv.groupby('rating').count()
```

	Name	year	movie_rated	length	genres	release_date	num_raters	num_review
rating								
8.4	1	1	1	1	1	1	1	
8.7	1	1	1	1	1	1	1	
8.8	2	2	2	2	2	2	2	
8.9	1	1	1	1	1	1	1	

```python
fiv.dtypes
```

```
Name             object
year              int64
movie_rated      object
length           object
genres           object
release_date     object
rating          float64
num_raters        int64
num_reviews       int64
dtype: object
```

```python
fiv.describe(include=object)
```

	Name	movie_rated	length	genres	release_date
count	5	5	5	5	5
unique	5	1	5	4	5
top	Inception	PG-13	2h 28min	Adventure; Drama; Fantasy;	16 July 2010 (USA)
freq	1	5	1	2	1

```
categorical_data=fiv.describe(include=object)
categorical_data
```

	Name	movie_rated	length	genres	release_date
count	5	5	5	5	5
unique	5	1	5	4	5
top	Inception	PG-13	2h 28min	Adventure; Drama; Fantasy;	16 July 2010 (USA)
freq	1	5	1	2	1

```
numerical_data=fiv.describe()
numerical_data
```

	year	rating	num_raters	num_reviews
count	5.000000	5.000000	5.000000e+00	5.000000
mean	2005.600000	8.720000	1.619043e+06	3680.800000
std	5.029911	0.192354	2.158349e+05	1073.180414
min	2001.000000	8.400000	1.440188e+06	2559.000000
25%	2002.000000	8.700000	1.470329e+06	2979.000000
50%	2003.000000	8.800000	1.593859e+06	3681.000000
75%	2010.000000	8.800000	1.609165e+06	3820.000000
max	2012.000000	8.900000	1.981675e+06	5365.000000

```
numerical_data.groupby('year').min()
```

Out[178...

	rating	num_raters	num_reviews
year			
5.000000	5.000000	5.000000e+00	5.000000
5.029911	0.192354	2.158349e+05	1073.180414
2001.000000	8.400000	1.440188e+06	2559.000000
2002.000000	8.700000	1.470329e+06	2979.000000
2003.000000	8.800000	1.593859e+06	3681.000000
2005.600000	8.720000	1.619043e+06	3680.800000
2010.000000	8.800000	1.609165e+06	3820.000000
2012.000000	8.900000	1.981675e+06	5365.000000

In [180...

```python
numerical_data.groupby('year').max()
```

Out[180...

	rating	num_raters	num_reviews
year			
5.000000	5.000000	5.000000e+00	5.000000
5.029911	0.192354	2.158349e+05	1073.180414
2001.000000	8.400000	1.440188e+06	2559.000000
2002.000000	8.700000	1.470329e+06	2979.000000
2003.000000	8.800000	1.593859e+06	3681.000000
2005.600000	8.720000	1.619043e+06	3680.800000
2010.000000	8.800000	1.609165e+06	3820.000000
2012.000000	8.900000	1.981675e+06	5365.000000

In [182...

```python
fiv
```

	Name	year	movie_rated	length	genres	release_date	rating	num_raters
0	Inception	2010	PG-13	2h 28min	Action; Adventure; Sci-Fi;	16 July 2010 (USA)	8.8	1981675
1	The Lord of the Rings: The Fellowship of the Ring	2001	PG-13	2h 58min	Action; Adventure; Drama;	19 December 2001 (USA)	8.8	1609165
2	The Lord of the Rings: The Return of the King	2003	PG-13	3h 21min	Adventure; Drama; Fantasy;	17 December 2003 (USA)	8.9	1593859
3	The Dark Knight Rises	2012	PG-13	2h 44min	Action; Adventure;	20 July 2012 (USA)	8.4	1470329
4	The Lord of the Rings: The Two Towers	2002	PG-13	2h 59min	Adventure; Drama; Fantasy;	18 December 2002 (USA)	8.7	1440188

```python
fiv.groupby('movie_rated')['rating'].mean()
```

```
movie_rated
PG-13    8.72
Name: rating, dtype: float64
```

```python
df.groupby('movie_rated')['rating'].median()
```

```
movie_rated
Approved     8.30
G            8.00
GP           7.75
M/PG         7.70
NC-17        7.50
Not Rated    8.00
Null         8.90
PG           7.70
PG-13        7.30
Passed       8.10
R            7.60
TV-14        7.40
TV-MA        7.00
TV-PG        6.30
Name: rating, dtype: float64
```

```python
df.groupby('movie_rated')['rating'].mean()
```

```
movie_rated
Approved      8.310000
G             7.809091
GP            7.750000
M/PG          7.700000
NC-17         7.500000
Not Rated     7.732258
Null          8.900000
PG            7.545918
PG-13         7.269216
Passed        8.033333
R             7.557732
TV-14         7.400000
TV-MA         6.833333
TV-PG         6.300000
Name: rating, dtype: float64
```

```python
df.groupby(['movie_rated'])[['num_raters','num_reviews']].mean()
```

movie_rated	num_raters	num_reviews
Approved	291764.800000	675.800000
G	472524.236364	636.909091
GP	67977.500000	248.000000
M/PG	50313.000000	270.000000
NC-17	180957.000000	833.000000
Not Rated	88552.354839	322.967742
Null	103674.000000	146.000000
PG	368056.158163	723.433673
PG-13	450626.919608	1196.815686
Passed	151585.000000	438.222222
R	415007.945508	1005.552283
TV-14	30652.000000	258.000000
TV-MA	74451.333333	343.333333
TV-PG	21976.000000	366.000000

```python
df.isnull().sum()
```

```
Name              0
year              0
movie_rated       0
length            0
genres            0
release_date      0
rating            0
num_raters        0
num_reviews       0
dtype: int64
```

```python
fiv.pivot(columns='num_reviews').plot()
```

<Axes: >
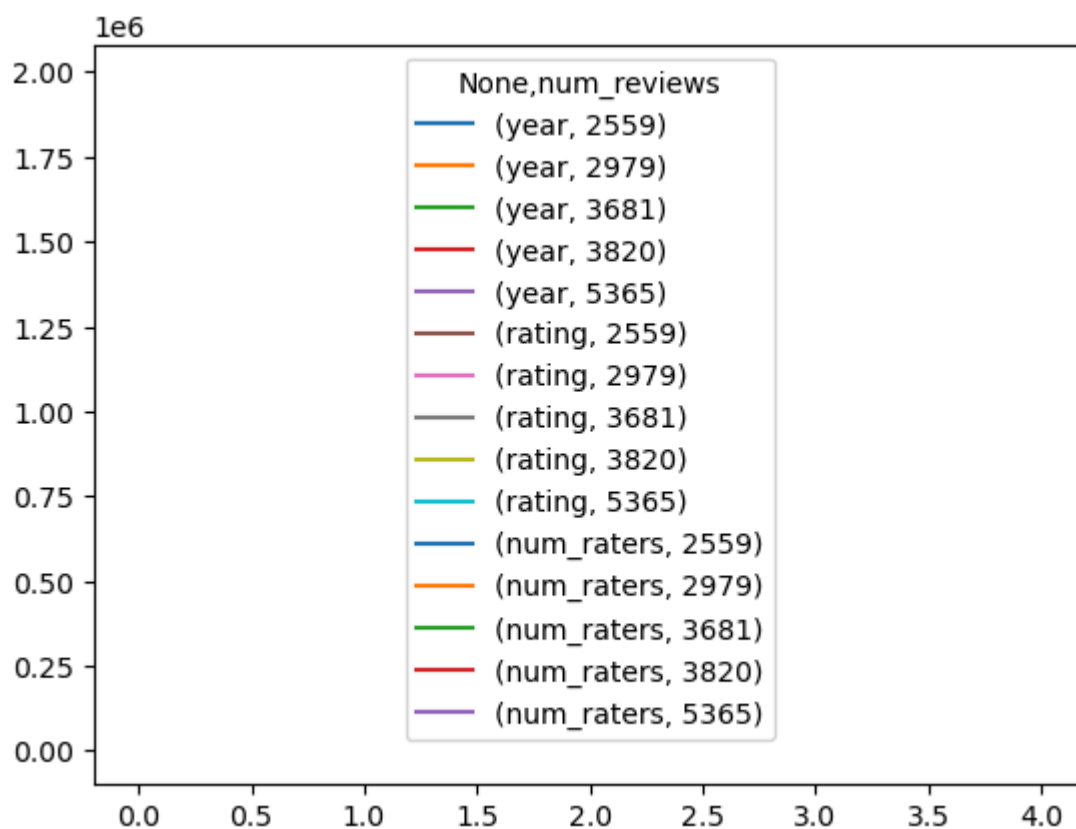
```python
df
```

Out[236...

	Name	year	movie_rated	length	genres	release_date	rating	num_rate
0	Inception	2010	PG-13	2h 28min	Action; Adventure; Sci-Fi;	16 July 2010 (USA)	8.8	198167
1	The Lord of the Rings: The Fellowship of the Ring	2001	PG-13	2h 58min	Action; Adventure; Drama;	19 December 2001 (USA)	8.8	160916
2	The Lord of the Rings: The Return of the King	2003	PG-13	3h 21min	Adventure; Drama; Fantasy;	17 December 2003 (USA)	8.9	159385
3	The Dark Knight Rises	2012	PG-13	2h 44min	Action; Adventure;	20 July 2012 (USA)	8.4	147032
4	The Lord of the Rings: The Two Towers	2002	PG-13	2h 59min	Adventure; Drama; Fantasy;	18 December 2002 (USA)	8.7	144018
...	...	...	...	...	...	...	...	
1495	Catch-22	1970	R	2h 2min	Comedy; Drama; War;	24 June 1970 (USA)	7.1	2142
1496	The Great Raid	2005	R	2h 12min	Action; Drama; War;	12 August 2005 (USA)	6.7	2096
1497	Saints and Soldiers	2003	PG-13	1h 30min	Action; Drama; War;	25 March 2005 (USA)	6.7	1973
1498	Stop-Loss	2008	R	1h 52min	Drama; War;	28 March 2008 (USA)	6.4	1945
1499	The Alamo	2004	PG-13	2h 17min	Drama; History; War;	9 April 2004 (USA)	6.0	1929

1500 rows × 9 columns

In [242... `del df["release_date"]`

In [244... `df`

	Name	year	movie_rated	length	genres	rating	num_raters	num_reviev
0	Inception	2010	PG-13	2h 28min	Action; Adventure; Sci-Fi;	8.8	1981675	38
1	The Lord of the Rings: The Fellowship of the Ring	2001	PG-13	2h 58min	Action; Adventure; Drama;	8.8	1609165	53
2	The Lord of the Rings: The Return of the King	2003	PG-13	3h 21min	Adventure; Drama; Fantasy;	8.9	1593859	36
3	The Dark Knight Rises	2012	PG-13	2h 44min	Action; Adventure;	8.4	1470329	29
4	The Lord of the Rings: The Two Towers	2002	PG-13	2h 59min	Adventure; Drama; Fantasy;	8.7	1440188	25
...	...	...	...	...	...	...	...	
1495	Catch-22	1970	R	2h 2min	Comedy; Drama; War;	7.1	21424	1
1496	The Great Raid	2005	R	2h 12min	Action; Drama; War;	6.7	20965	1
1497	Saints and Soldiers	2003	PG-13	1h 30min	Action; Drama; War;	6.7	19730	1
1498	Stop-Loss	2008	R	1h 52min	Drama; War;	6.4	19456	1
1499	The Alamo	2004	PG-13	2h 17min	Drama; History; War;	6.0	19290	3

1500 rows × 8 columns

◀ ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ ▶

```python
Filter1=df["rating"]>8
```

```python
Filter2=df['rating']>8.5
```

```python
num_of_TopRatings=df[Filter2].count()
```

```
In [280...  print(num_of_TopRatings)

Name            61
year            61
movie_rated     61
length          61
genres          61
rating          61
num_raters      61
num_reviews     61
dtype: int64
```
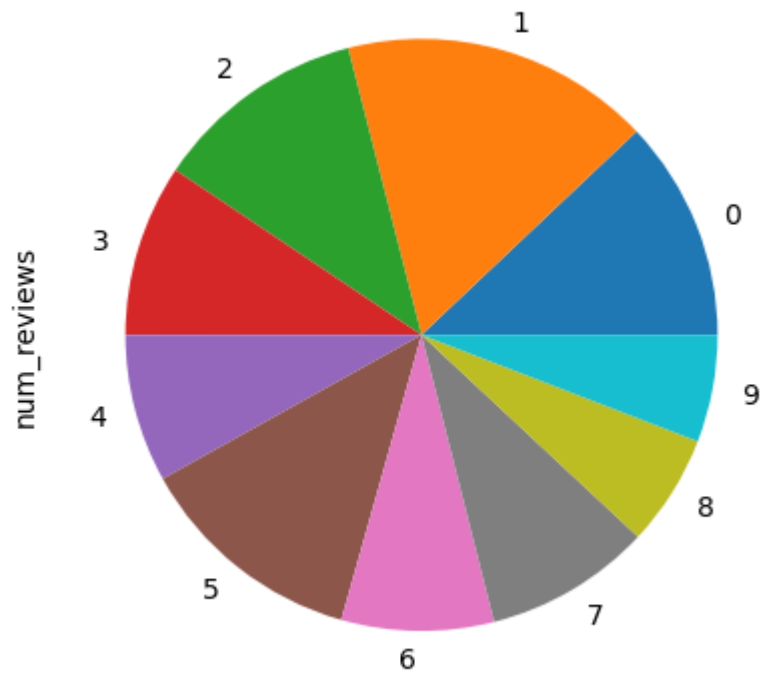
Data Visualization

```
In [291...  import matplotlib.pyplot as plt
           import seaborn as sns
```

```
In [299...  vis1=fiv['num_reviews'].plot(kind='bar',figsize=(10,5))
```
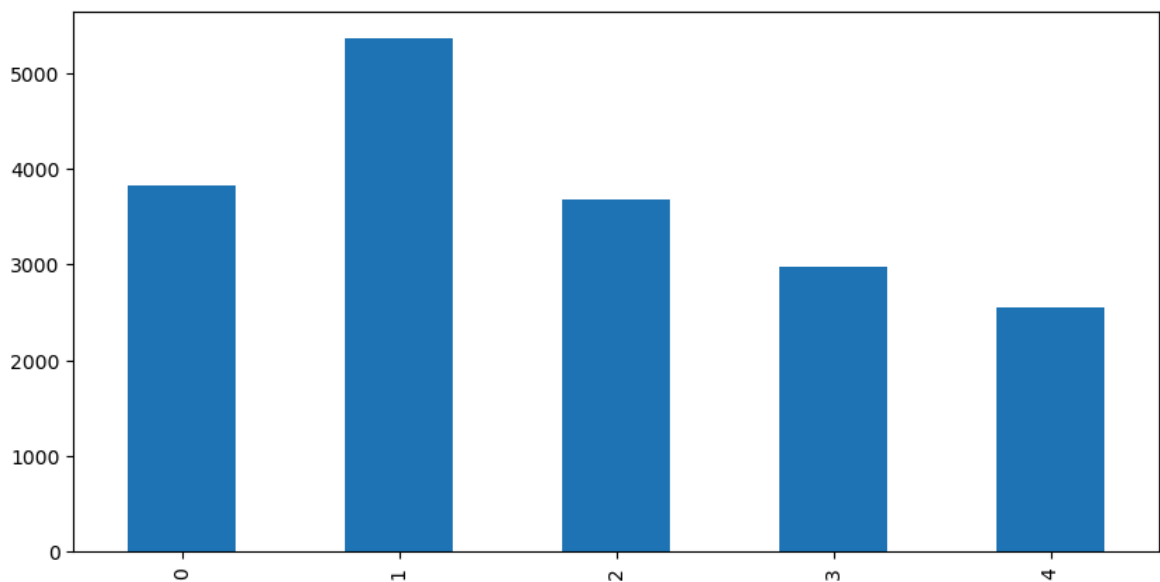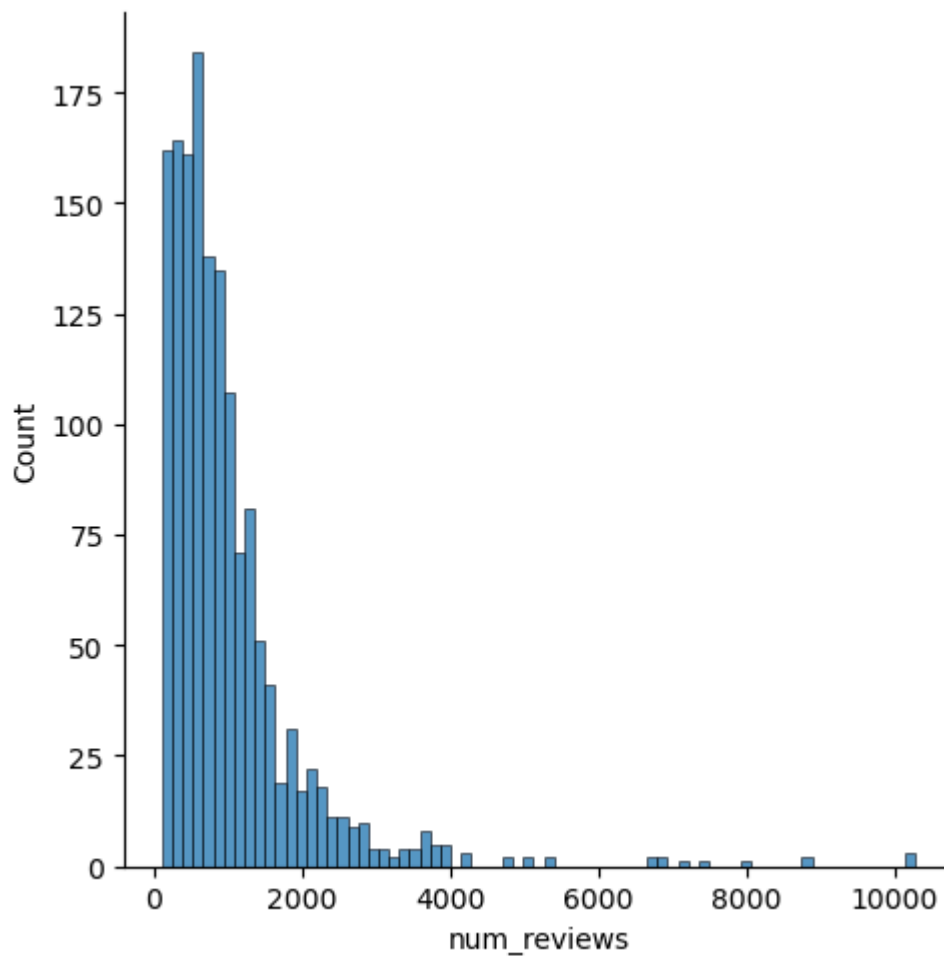


```
In [360...  v=df['num_reviews'].head(10).plot(kind='pie')
```
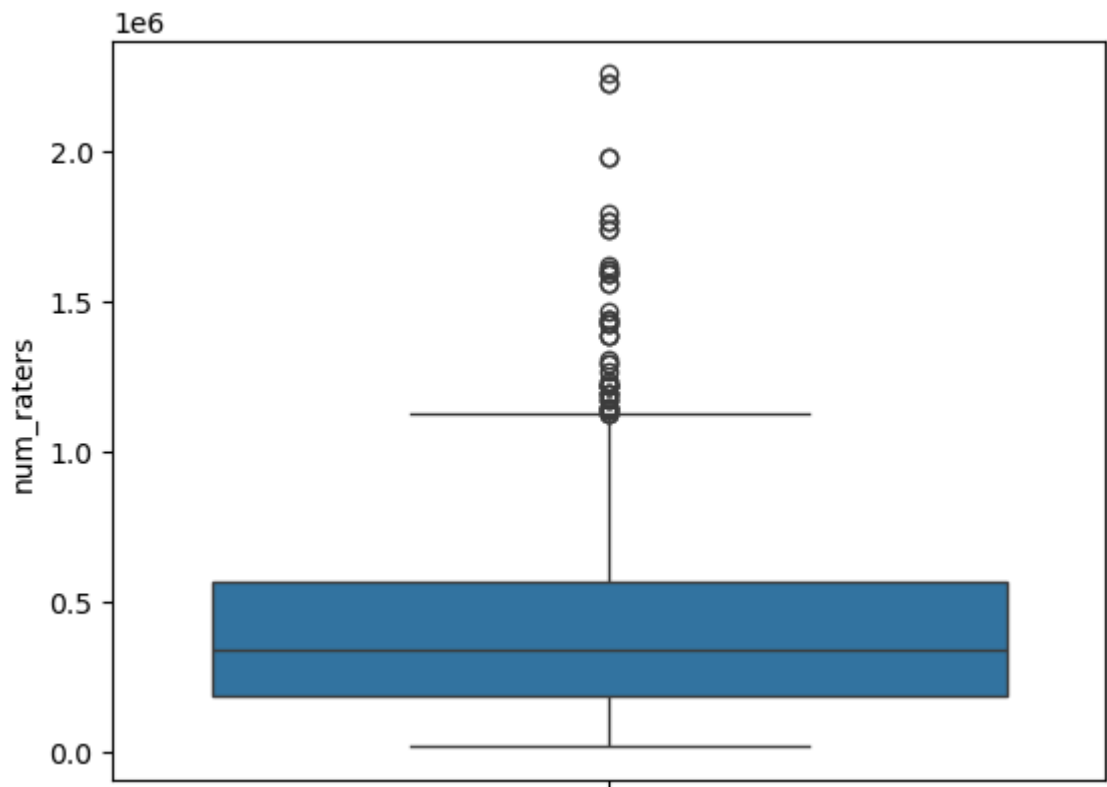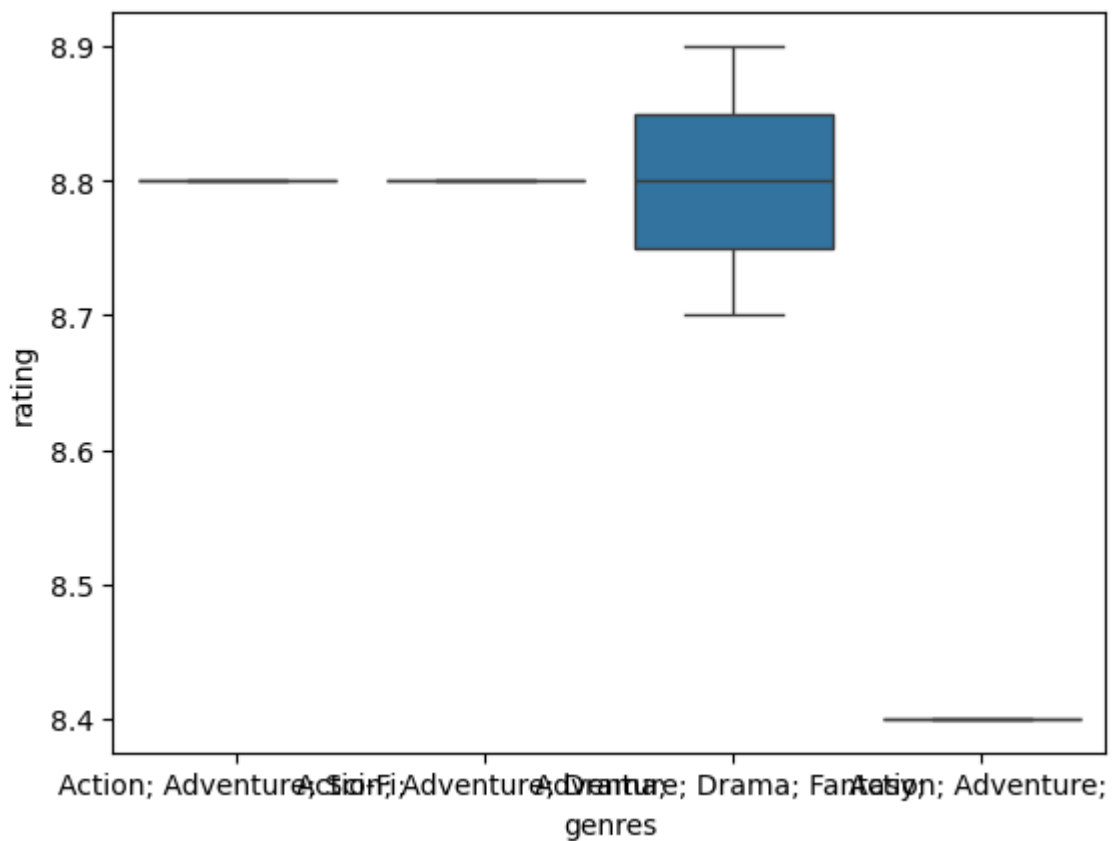
```
In [350...  plt.show(v)
```

```
In [301...  vis2=sns.displot(df["num_reviews"])
```



```
In [303...  vis3=sns.boxplot(df["num_raters"])
```
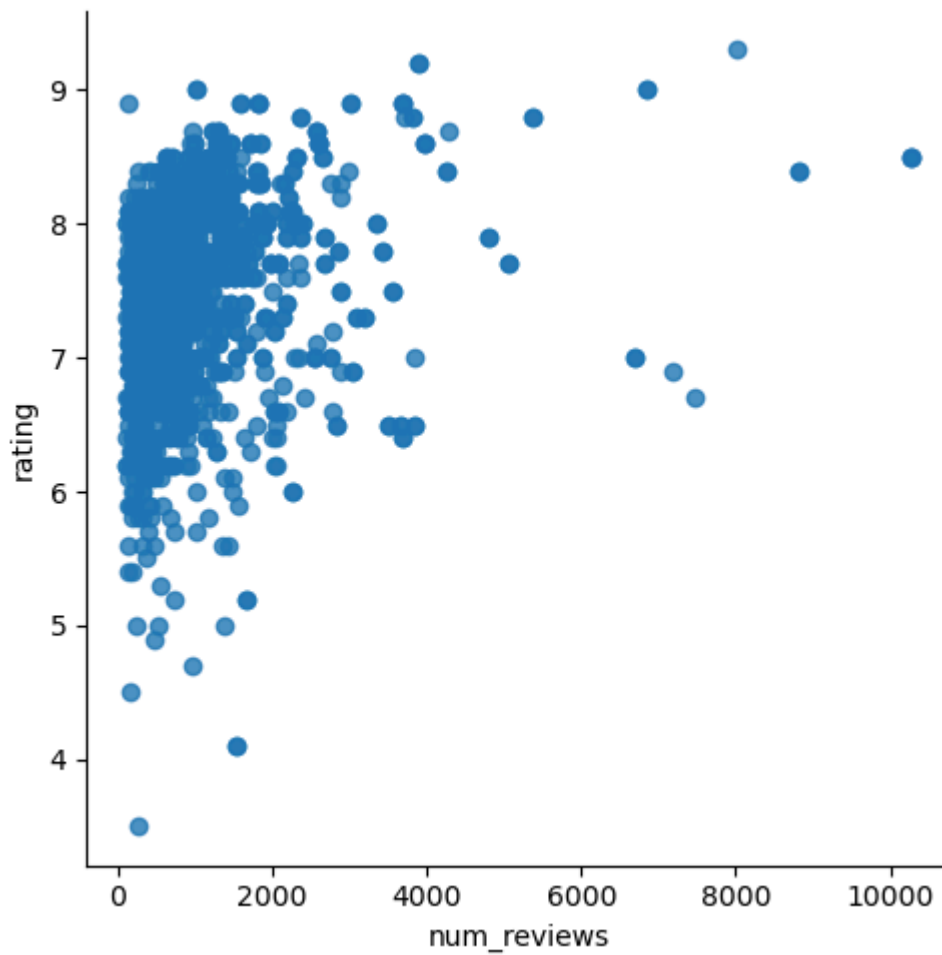
In [307… `vis4=sns.boxplot(data=fiv,x='genres',y='rating')`
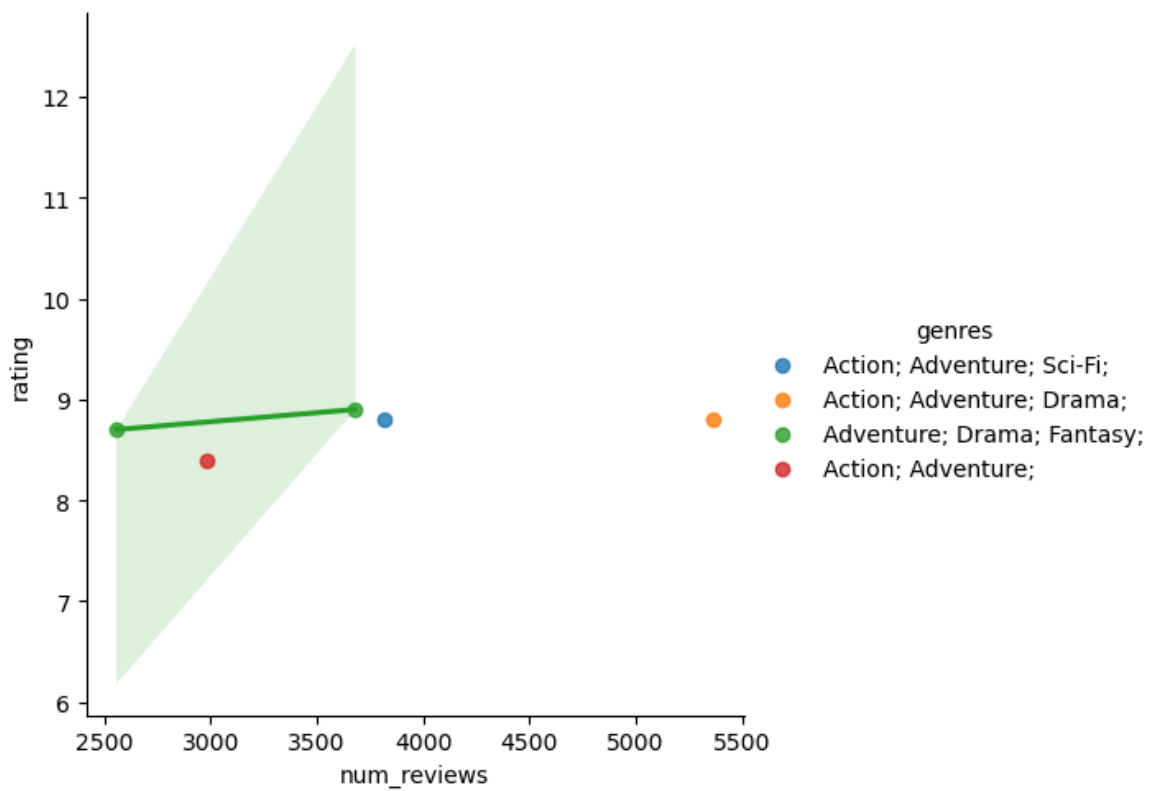

In [313… `vis5=sns.lmplot(data=df,x='num_raters',y='rating')`
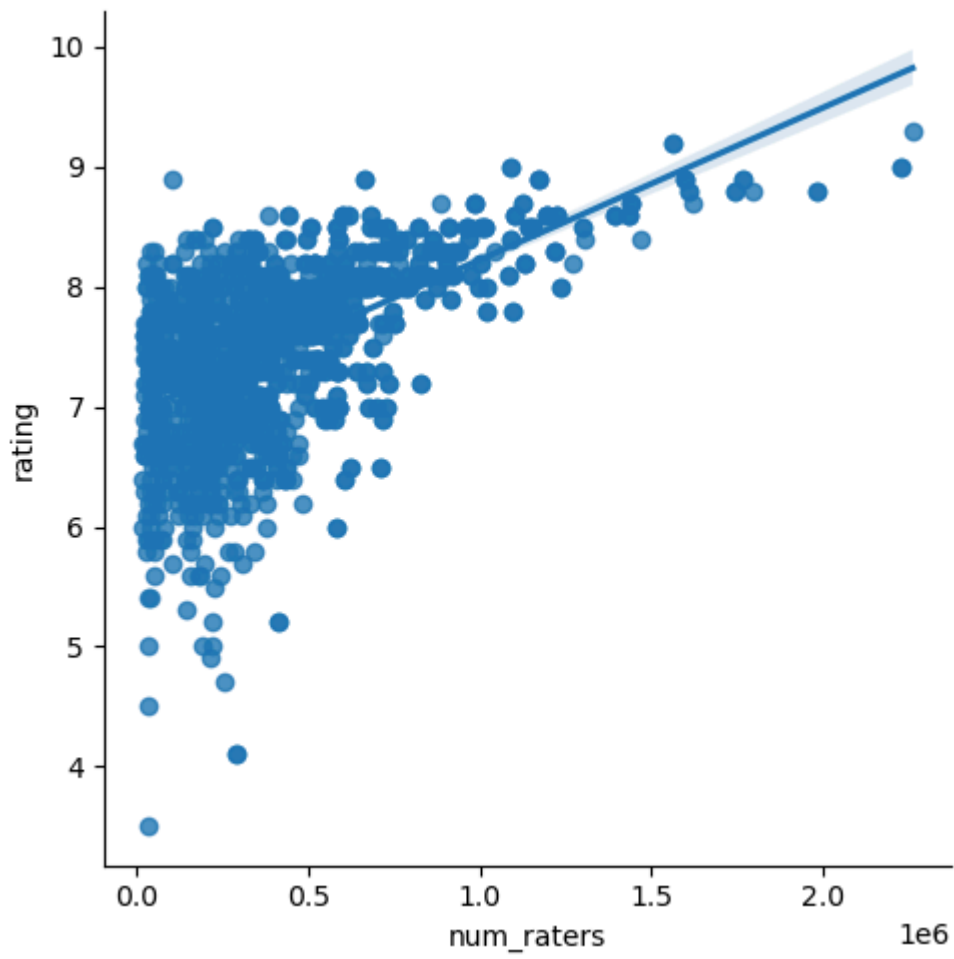
```
vis5=sns.lmplot(data=df,x="num_reviews",y='rating',fit_reg=False)
```

```
vis6=sns.lmplot(data=fiv,x="num_reviews",y="rating",hue='genres')
plt.show(vis6)
```



```
plt.show(vis6)
```