

```
In [2]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
In [3]: df = pd.read_csv(r"C:\Users\sharm\Downloads\amazon_prime.csv")
```

In [4]: df

Out[4]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Di
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	S
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	1
...	...	...	...	...	...	...	...	...	
7784	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car...	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	TV-MA	
7785	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	TV-14	
7786	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	TV-MA	
7787	s7786	TV Show	Zumbo's Just Desserts	NaN	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	TV-PG	s
7788	s7787	Movie	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	Sam Dunn	NaN	United Kingdom, Canada, United States	March 1, 2020	TV-MA	

7789 rows × 11 columns

In [5]: df.head(5)

Out[5]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 min
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	80 min
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min

In [6]: `df.tail()`

Out[6]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration
7784	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car...	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	TV-MA	96
7785	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	TV-14	114
7786	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	TV-MA	44
7787	s7786	TV Show	Zumbo's Just Desserts	NaN	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	TV-PG	Season 1
7788	s7787	Movie	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	Sam Dunn	NaN	United Kingdom, Canada, United States	March 1, 2020	TV-MA	90

In [7]: `df.shape`

Out[7]: (7789, 11)

In [8]: `df.size`

Out[8]: 85679

In [9]: `df.dtypes`

Out[9]:

Show_Id	object
Category	object
Title	object
Director	object
Cast	object
Country	object
Release_Date	object
Rating	object
Duration	object
Type	object
Description	object
dtype:	object

In [10]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7789 entries, 0 to 7788
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Show_Id         7789 non-null   object
1   Category        7789 non-null   object
2   Title           7789 non-null   object
3   Director        5401 non-null   object
4   Cast            7071 non-null   object
5   Country         7282 non-null   object
6   Release_Date    7779 non-null   object
7   Rating          7782 non-null   object
8   Duration        7789 non-null   object
9   Type            7789 non-null   object
10  Description      7789 non-null   object
dtypes: object(11)
memory usage: 669.5+ KB
```

In [11]: df.describe()

Out[11]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating
count	7789	7789	7789	5401	7071	7282	7779	7782
unique	7787	2	7787	4050	6831	681	1565	14
top	s6621	Movie	The Lost Okoroshi	Raúl Campos, Jan Suter	David Attenborough	United States	January 1, 2020	TV-MA
freq	2	5379	2	18	18	2556	118	2865

In [12]: df[df.duplicated()]

Out[12]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration
6300	s684	Movie	Backfire	Dave Patten	Black Deniro, Byron "Squally" Vinson, Dominic ...	United States	April 5, 2019	TV-MA	97 min
6622	s6621	Movie	The Lost Okoroshi	Abba T. Makama	Seun Ajayi, Judith Audu, Tope Tedela, Ifu Enna...	Nigeria	September 4, 2020	TV-MA	94 min

```
In [13]: df.drop_duplicates(inplace=True)
```

```
In [14]: df.shape
```

```
Out[14]: (7787, 11)
```

```
In [15]: df.isnull().sum()
```

```
Out[15]: Show_Id      0
          Category    0
          Title       0
          Director   2388
          Cast       718
          Country    507
          Release_Date 10
          Rating      7
          Duration    0
          Type        0
          Description 0
          dtype: int64
```

```
In [16]: df.fillna({'Director':'unavailable','Cast':'unavailable','Country':'unavailable'})
df.isna().sum()
```

```
Out[16]: Show_Id      0
          Category    0
          Title       0
          Director    0
          Cast        0
          Country     0
          Release_Date 10
          Rating      0
          Duration    0
          Type        0
          Description 0
          dtype: int64
```

```
In [17]: df.dtypes
```

```
Out[17]: Show_Id      object
          Category    object
          Title       object
          Director    object
          Cast        object
          Country     object
          Release_Date object
          Rating      object
          Duration    object
          Type        object
          Description object
          dtype: object
```

```
In [18]: df['date'] = pd.to_datetime(df['Release_Date'])
```

In [19]: df



Out[19]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating
0	s1	TV Show	3%	unavailable	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13
...	...	...	...	...	...	...	...	...
7784	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car...	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	TV-MA
7785	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	TV-14
7786	s7785	Movie	Zulu Man in Japan	unavailable	Nasty C	unavailable	September 25, 2020	TV-MA
7787	s7786	TV Show	Zumbo's Just Desserts	unavailable	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	TV-PG
7788	s7787	Movie	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	Sam Dunn	unavailable	United Kingdom, Canada, United States	March 1, 2020	TV-MA

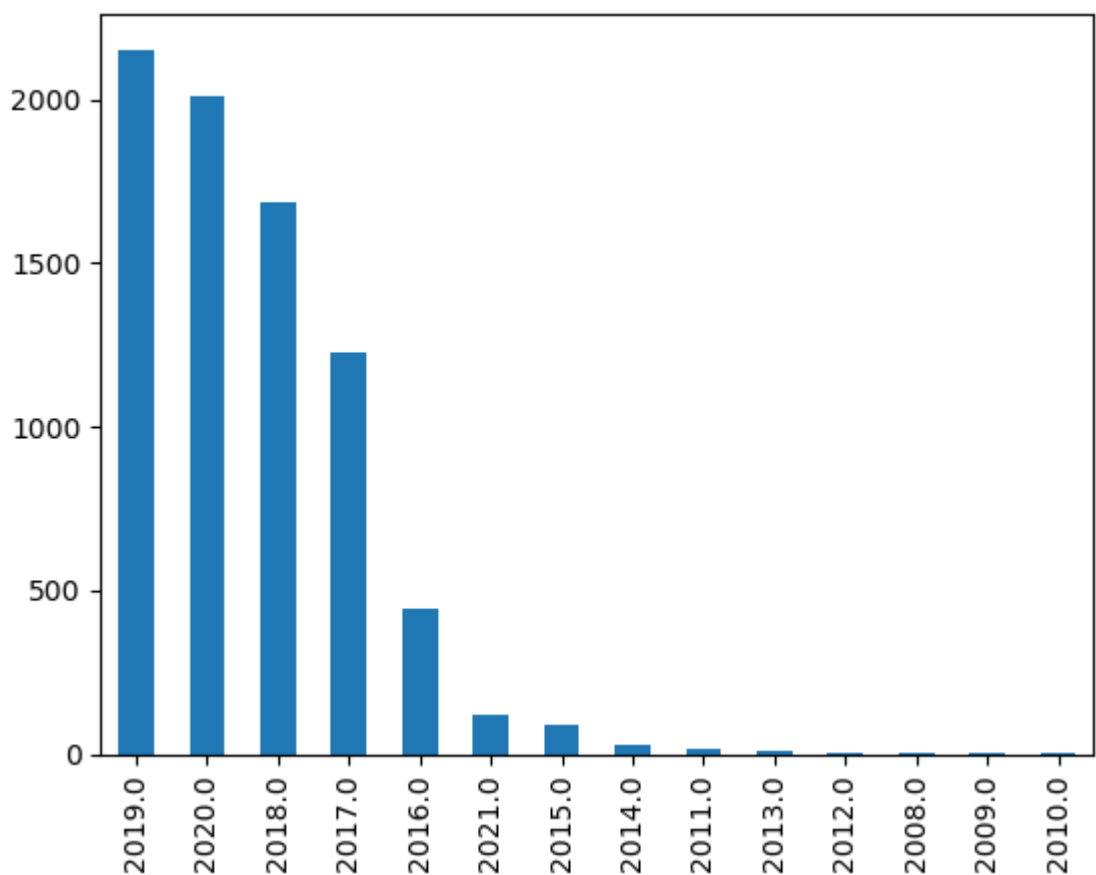
7787 rows × 12 columns

```
In [20]: df['date'].dt.year.value_counts()
```

```
Out[20]: 2019.0    2153
         2020.0    2009
         2018.0    1685
         2017.0    1225
         2016.0     443
         2021.0     117
         2015.0      88
         2014.0      25
         2011.0      13
         2013.0      11
         2012.0       3
         2008.0       2
         2009.0       2
         2010.0       1
         Name: date, dtype: int64
```

```
In [21]: df['date'].dt.year.value_counts().plot(kind='bar')
```

```
Out[21]: <Axes: >
```



```
In [23]: df[df.date.isnull()]
```

Out[23]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Ratin
258	s259	TV Show	A Young Doctor's Notebook and Other Stories	unavailable	Daniel Radcliffe, Jon Hamm, Adam Godley, Chris...	United Kingdom	NaN	TV-M
549	s550	TV Show	Anthony Bourdain: Parts Unknown	unavailable	Anthony Bourdain	United States	NaN	TV-P
2263	s2264	TV Show	Frasier	unavailable	Kelsey Grammer, Jane Leeves, David Hyde Pierce...	United States	NaN	TV-P
2288	s2289	TV Show	Friends	unavailable	Jennifer Aniston, Courteney Cox, Lisa Kudrow, ...	United States	NaN	TV-1
2555	s2556	TV Show	Gunslinger Girl	unavailable	Yuuka Nanri, Kanako Mitsuhashi, Eri Sendai, Am...	Japan	NaN	TV-1
3374	s3375	TV Show	Kikoriki	unavailable	Igor Dmitriev	unavailable	NaN	TV-
3492	s3493	TV Show	La Familia P. Luche	unavailable	Eugenio Derbez, Consuelo Duval, Luis Manuel Áv...	United States	NaN	TV-1
3946	s3947	TV Show	Maron	unavailable	Marc Maron, Judd Hirsch, Josh Brener, Nora Zeh...	United States	NaN	TV-M
5138	s5138	TV Show	Red vs. Blue	unavailable	Burnie Burns, Jason Saldaña, Gustavo Sorola, G...	United States	NaN	N
6066	s6066	TV Show	The Adventures of Figaro Pho	unavailable	Luke Jurevicius, Craig Behenna, Charlotte Haml...	Australia	NaN	TV-Y

```
In [24]: recent_date=df['date'].max()
df.fillna({'date': recent_date},inplace= True)
```

C:\Users\sharm\AppData\Local\Temp\ipykernel\_19236\1330307752.py:2: DeprecationWarning: In a future version, `df.iloc[:, i] = newvals` will attempt to set the values inplace instead of always setting a new array. To retain the old behavior, use either `df[df.columns[i]] = newvals` or, if columns are non-unique, `df.isetitem(i, newvals)`

```
df.fillna({'date': recent_date},inplace= True)
```

```
In [25]: df[df.Show_Id == 's259']
```

Out[25]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Dur
258	s259	TV Show	A Young Doctor's Notebook and Other Stories	unavailable	Daniel Radcliffe, Jon Hamm, Adam Godley, Chris...	United Kingdom	NaN	TV-MA	Se

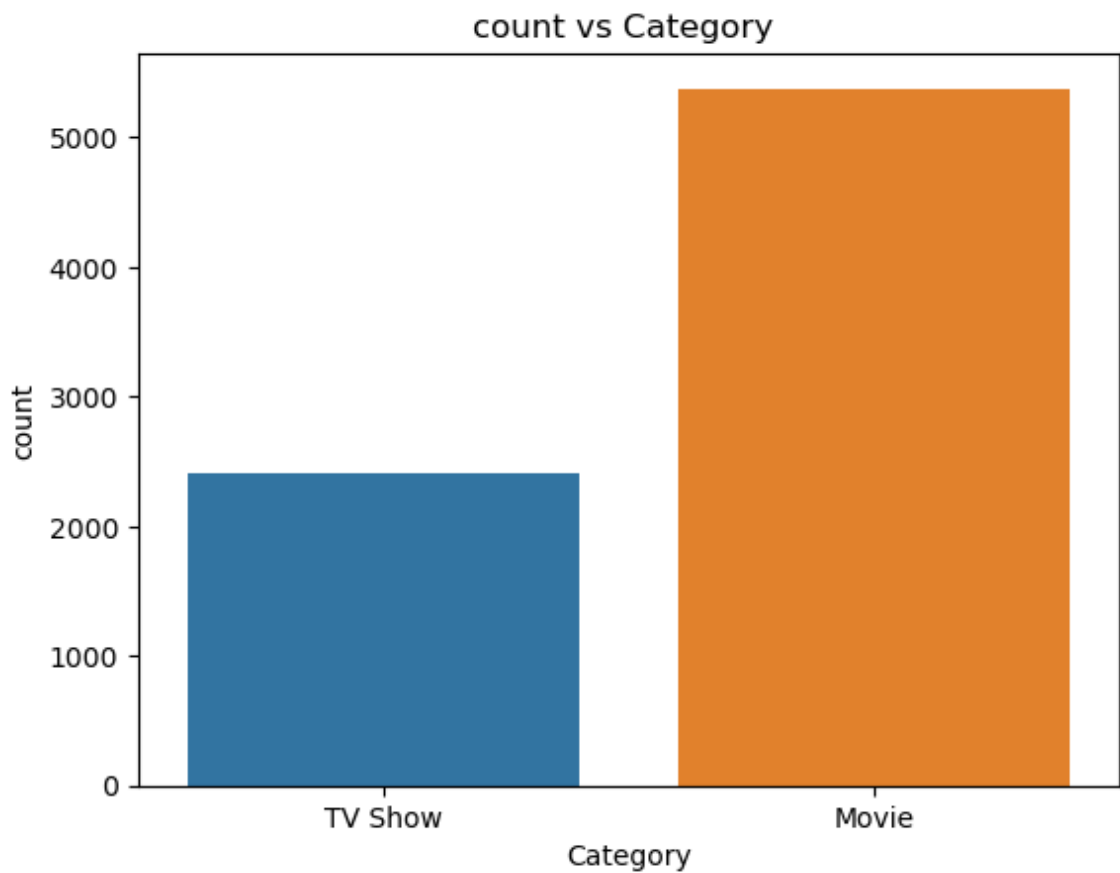


```
In [26]: df.Type.value_counts()
```

Out[26]: Documentaries  
334  
Stand-Up Comedy  
321  
Dramas, International Movies  
320  
Comedies, Dramas, International Movies  
243  
Dramas, Independent Movies, International Movies  
215  
  
...  
Crime TV Shows, Romantic TV Shows, Spanish-Language TV Shows  
1  
Crime TV Shows, International TV Shows, Reality TV  
1  
International Movies, Romantic Movies, Sci-Fi & Fantasy  
1  
Stand-Up Comedy & Talk Shows, TV Mysteries, TV Sci-Fi & Fantasy  
1  
International TV Shows, Spanish-Language TV Shows, Stand-Up Comedy & Talk Shows  
1  
Name: Type, Length: 492, dtype: int64

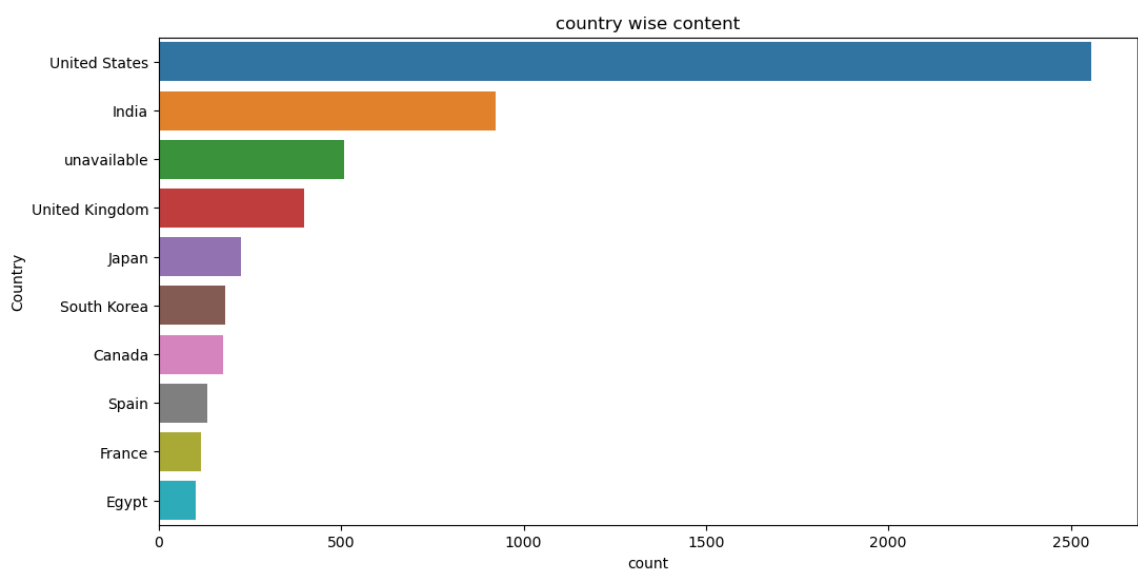
```
In [27]: sns.countplot(x = 'Category', data = df)
plt.title('count vs Category ')
```

```
Out[27]: Text(0.5, 1.0, 'count vs Category ')
```



```
In [28]: plt.figure(figsize = (12,6))
sns.countplot(y = 'Country',order = df['Country'].value_counts().index[0:10])
plt.title('country wise content')
```

```
Out[28]: Text(0.5, 1.0, 'country wise content')
```

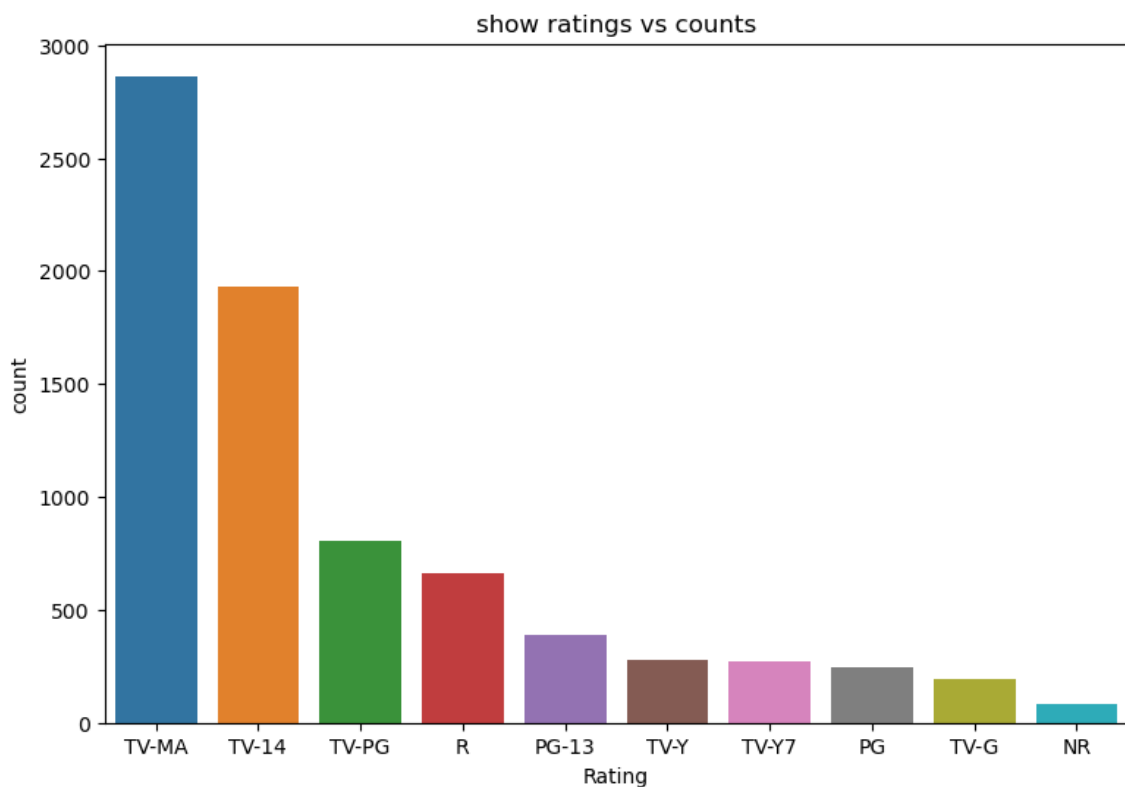


```
In [29]: df.Rating.value_counts()
```

```
Out[29]: TV-MA      2863
TV-14      1931
TV-PG      806
R          665
PG-13      386
TV-Y       280
TV-Y7      271
PG         247
TV-G       194
NR         84
G          39
unavailable 7
TV-Y7-FV    6
UR          5
NC-17       3
Name: Rating, dtype: int64
```

```
In [30]: plt.figure(figsize = (9,6))
sns.countplot( x = 'Rating', order = df['Rating'].value_counts().index[0:10]
plt.title('show ratings vs counts')
```

```
Out[30]: Text(0.5, 1.0, 'show ratings vs counts')
```



```
In [31]: plt.figure(figsize=(12,9))  
sns.countplot(y = 'Type',order = df['Type'].value_counts().index[0:20],data  
plt.title('top genres')
```

```
Out[31]: Text(0.5, 1.0, 'top genres')
```

