

```
In [49]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
data = pd.read_csv("D:\\Software Very Important\\Data Analysis\\netflix_dataset.csv")
data
```

Out[49]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA
...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	F
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	F
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy	United States	January 11, 2020	2006	PG

	show_id	type	title	director	cast	country	date_added	release_year	rating
					Chase, Kate Ma...				
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah- Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14

In [50]: `data.shape # gives rows and columns`

Out[50]: (8807, 12)

In [51]: `data.info() # Tells you the count and data types of your data. (Objects is str)`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   show_id          8807 non-null   object 
 1   type              8807 non-null   object 
 2   title             8807 non-null   object 
 3   director          6173 non-null   object 
 4   cast               7982 non-null   object 
 5   country            7976 non-null   object 
 6   date_added        8797 non-null   object 
 7   release_year      8807 non-null   int64  
 8   rating             8803 non-null   object 
 9   duration            8804 non-null   object 
 10  listed_in          8807 non-null   object 
 11  description         8807 non-null   object 
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
```

In [52]: `data.describe() # Tells you some basic statistics about your data`

Out[52]: `release_year`

count	8807.000000
mean	2014.180198
std	8.819312
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

In [53]: `data.isna() # False is not null value, True is null value`

Out[53]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	
0	False	False	False	False	True	False	False	False	False	False	
1	False	False	False	True	False	False	False	False	False	False	
2	False	False	False	False	False	True	False	False	False	False	
3	False	False	False	True	True	True	False	False	False	False	
4	False	False	False	True	False	False	False	False	False	False	
...
8802	False	False	False	False	False	False	False	False	False	False	
8803	False	False	False	True	True	True	False	False	False	False	
8804	False	False	False	False	False	False	False	False	False	False	
8805	False	False	False	False	False	False	False	False	False	False	
8806	False	False	False	False	False	False	False	False	False	False	

8807 rows × 12 columns



In [54]: `data.isna().sum()`

Out[54]:

show_id	0
type	0
title	0
director	2634
cast	825
country	831
date_added	10
release_year	0
rating	4
duration	3
listed_in	0
description	0
dtype: int64	

In [55]: `data['date_added']=pd.to_datetime(data['date_added']) # Changing object to datetime`

In [56]: `data`

	show_id	type	title	director	cast	country	date_added	release_year	rating	
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV-MA	
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	2021-09-24	2021	TV-MA	
3	s4	TV Show	Jailbirds	NaN	NaN	NaN	2021-09-24	2021	TV-MA	
4	s5	TV Show	New Orleans	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021-09-24	2021	TV-MA
...	
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	2019-11-20	2007	F	
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2018	TV-Y7	
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	2019-11-01	2009	F	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy	United States	2020-01-11	2006	PG	

show_id	type	title	director	cast	country	date_added	release_year	rating
				Chase, Kate Ma...				
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah- Jane Dias, Raaghav Chanan...	India	2019-03-02	2015 TV-14

In [57]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   show_id          8807 non-null   object 
 1   type              8807 non-null   object 
 2   title             8807 non-null   object 
 3   director          6173 non-null   object 
 4   cast               7982 non-null   object 
 5   country            7976 non-null   object 
 6   date_added        8797 non-null   datetime64[ns]
 7   release_year      8807 non-null   int64  
 8   rating             8803 non-null   object 
 9   duration           8804 non-null   object 
 10  listed_in         8807 non-null   object 
 11  description        8807 non-null   object 
dtypes: datetime64[ns](1), int64(1), object(10)
memory usage: 825.8+ KB
```

In [58]: `data.fillna({'director':'Unavailable','cast':'Unavailable','country':'Unavailable'})
data.isna().sum()`

```
Out[58]: show_id      0
type          0
title          0
director       0
cast           0
country         0
date_added     10
release_year    0
rating          0
duration        0
listed_in       0
description     0
dtype: int64
```

In [60]: `data[data.date_added.isnull()]`

Out[60]:		show_id	type	title	director	cast	country	date_added	release_year
	6066	s6067	TV Show	A Young Doctor's Notebook and Other Stories	Unavailable	Daniel Radcliffe, Jon Hamm, Adam Godley, Chris...	United Kingdom	NaT	2013
	6174	s6175	TV Show	Anthony Bourdain: Parts Unknown	Unavailable	Anthony Bourdain	United States	NaT	2018
	6795	s6796	TV Show	Frasier	Unavailable	Kelsey Grammer, Jane Leeves, David Hyde Pierce...	United States	NaT	2003
	6806	s6807	TV Show	Friends	Unavailable	Jennifer Aniston, Courteney Cox, Lisa Kudrow, ...	United States	NaT	2003
	6901	s6902	TV Show	Gunslinger Girl	Unavailable	Yuuka Nanri, Kanako Mitsuhashi, Eri Sendai, Am...	Japan	NaT	2008
	7196	s7197	TV Show	Kikoriki	Unavailable	Igor Dmitriev	Unavailable	NaT	2010
	7254	s7255	TV Show	La Familia P. Luche	Unavailable	Eugenio Derbez, Consuelo Duval, Luis Manuel Áv...	United States	NaT	2012
	7406	s7407	TV Show	Maron	Unavailable	Marc Maron, Judd Hirsch, Josh Brener, Nora Zeh...	United States	NaT	2016
	7847	s7848	TV Show	Red vs. Blue	Unavailable	Burnie Burns, Jason Saldaña, Gustavo Sorola, G...	United States	NaT	2015
	8182	s8183	TV Show	The Adventures of Figaro Pho	Unavailable	Luke Jurevicius, Craig Behenna,	Australia	NaT	2015

show_id	type	title	director	cast	country	date_added	release_year	rating
		Charlotte	Lionel					

```
In [63]: most_recent_entry_data=data['date_added'].max()
data.fillna({'date_added':most_recent_entry_data},inplace=True)
```

```
In [64]: data[data.show_id=='s6067']
```

Out[64]:

show_id	type	title	director	cast	country	date_added	release_year	rating	
6066	s6067	TV Show	A Young Doctor's Notebook and Other Stories	Unavailable	Daniel Radcliffe, Jon Hamm, Adam Godley, Chris...	United Kingdom	2021-09-25	2013	TV-MA

```
In [65]: data[data.duration.isnull()]
```

Out[65]:

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in

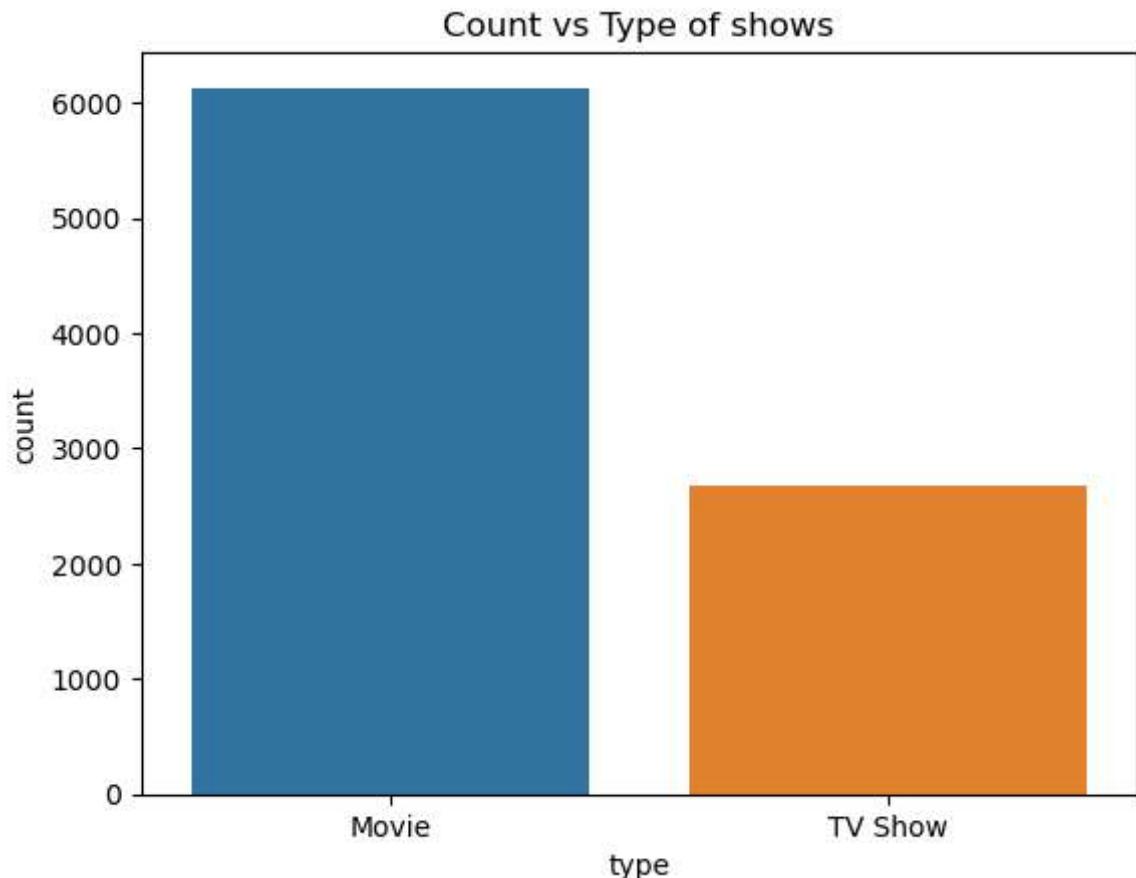
```
In [66]: data.type.value_counts()
```

Out[66]:

Movie	6131
TV Show	2676
Name: type, dtype: int64	

```
In [72]: sns.countplot(x='type',data=data)
plt.title('Count vs Type of shows')
```

Out[72]:

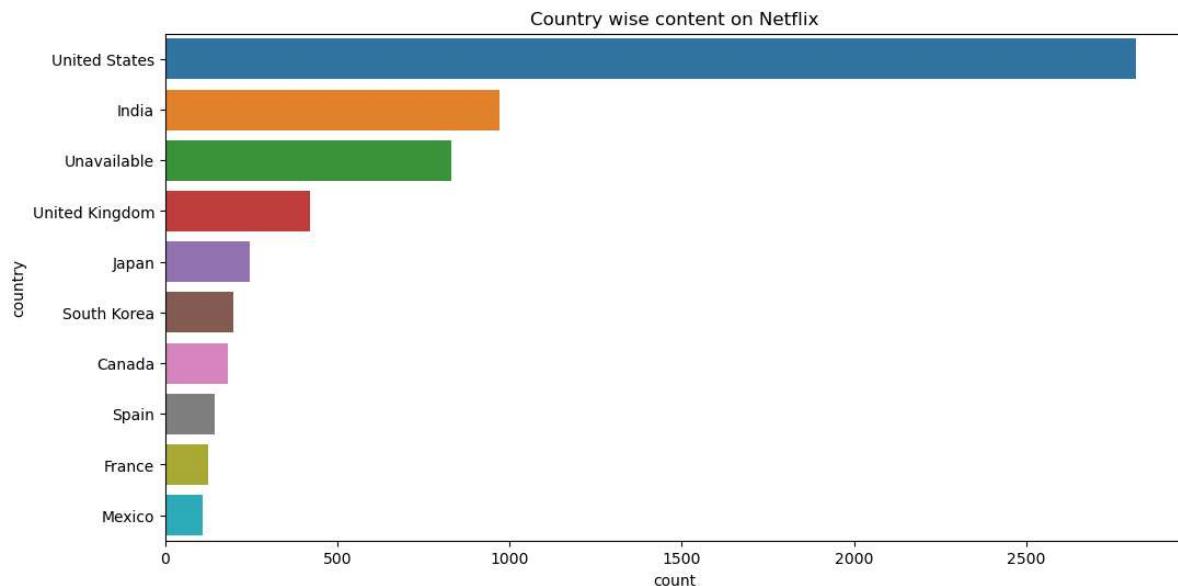


```
In [76]: data['country'].value_counts().head(10)
```

```
Out[76]: United States    2818
          India        972
          Unavailable   831
          United Kingdom 419
          Japan        245
          South Korea   199
          Canada       181
          Spain         145
          France       124
          Mexico       110
          Name: country, dtype: int64
```

```
In [77]: plt.figure(figsize=(12,6))
sns.countplot(y='country',order=data['country'].value_counts().index[0:10],data=data)
plt.title('Country wise content on Netflix')
```

```
Out[77]: Text(0.5, 1.0, 'Country wise content on Netflix')
```

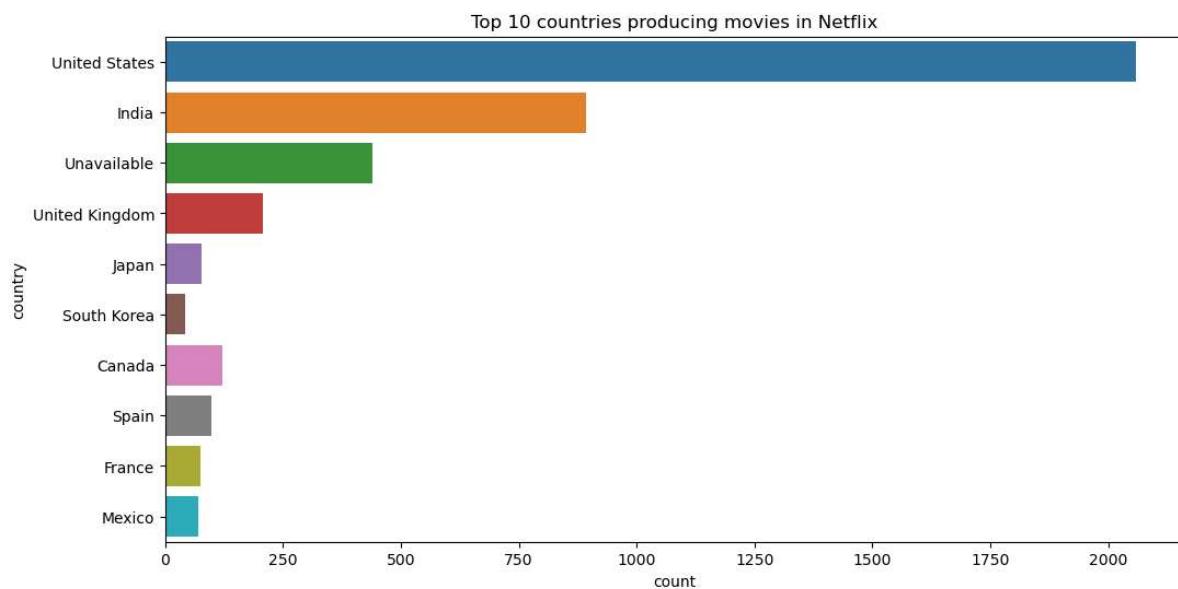


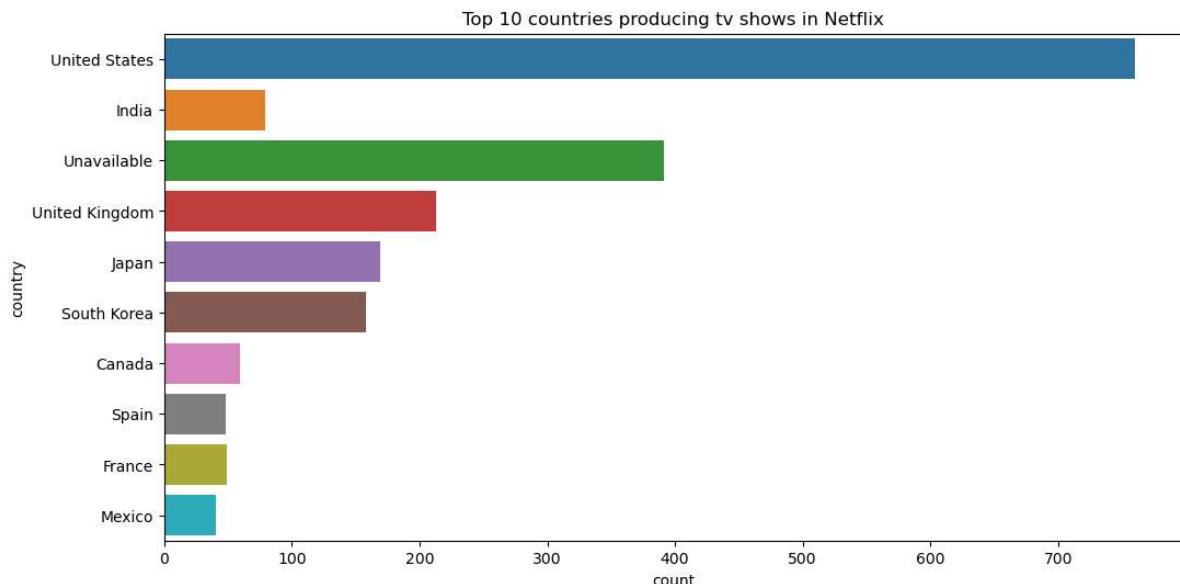
```
In [81]: movie_countries=data[data['type']=='Movie']
tv_show_countries=data[data['type']=='TV Show']
```

```
In [82]: plt.figure(figsize=(12,6))
sns.countplot(y='country',order=data['country'].value_counts().index[0:10],data=movie_countries)
plt.title("Top 10 countries producing movies in Netflix")

plt.figure(figsize=(12,6))
sns.countplot(y='country',order=data['country'].value_counts().index[0:10],data=tv_show_countries)
plt.title("Top 10 countries producing tv shows in Netflix")
```

Out[82]: Text(0.5, 1.0, 'Top 10 countries producing tv shows in Netflix')





```
In [83]: data.rating.value_counts()
```

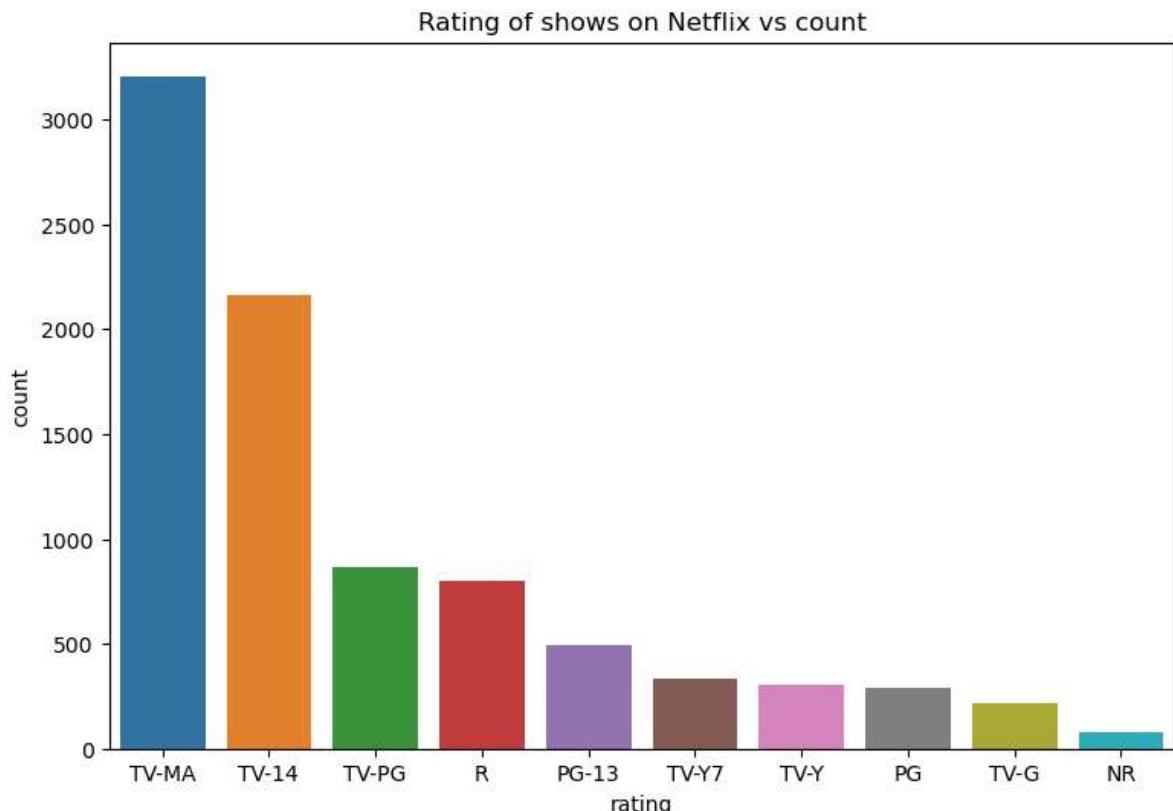
```
Out[83]:
```

rating	count
TV-MA	3207
TV-14	2160
TV-PG	863
R	799
PG-13	490
TV-Y7	334
TV-Y	307
PG	287
TV-G	220
NR	80
G	41
TV-Y7-FV	6
Unavailable	4
NC-17	3
UR	3
74 min	1
84 min	1
66 min	1

Name: rating, dtype: int64

```
In [84]: plt.figure(figsize=(9,6))
sns.countplot(x='rating',order=data['rating'].value_counts().index[0:10],data=data)
plt.title('Rating of shows on Netflix vs count')
```

```
Out[84]: Text(0.5, 1.0, 'Rating of shows on Netflix vs count')
```



```
In [85]: data.release_year.value_counts()[:20]
```

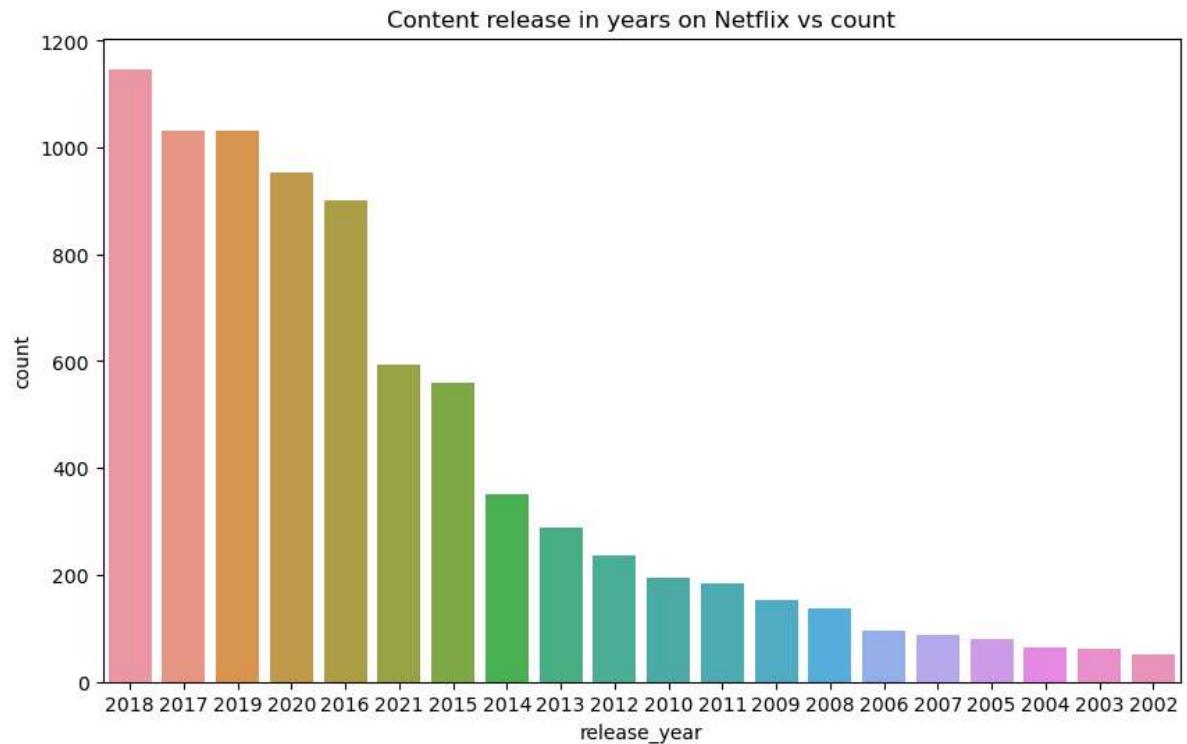
```
Out[85]:
```

release_year	count
2018	1147
2017	1032
2019	1030
2020	953
2016	902
2021	592
2015	560
2014	352
2013	288
2012	237
2010	194
2011	185
2009	152
2008	136
2006	96
2007	88
2005	80
2004	64
2003	61
2002	51

Name: release_year, dtype: int64

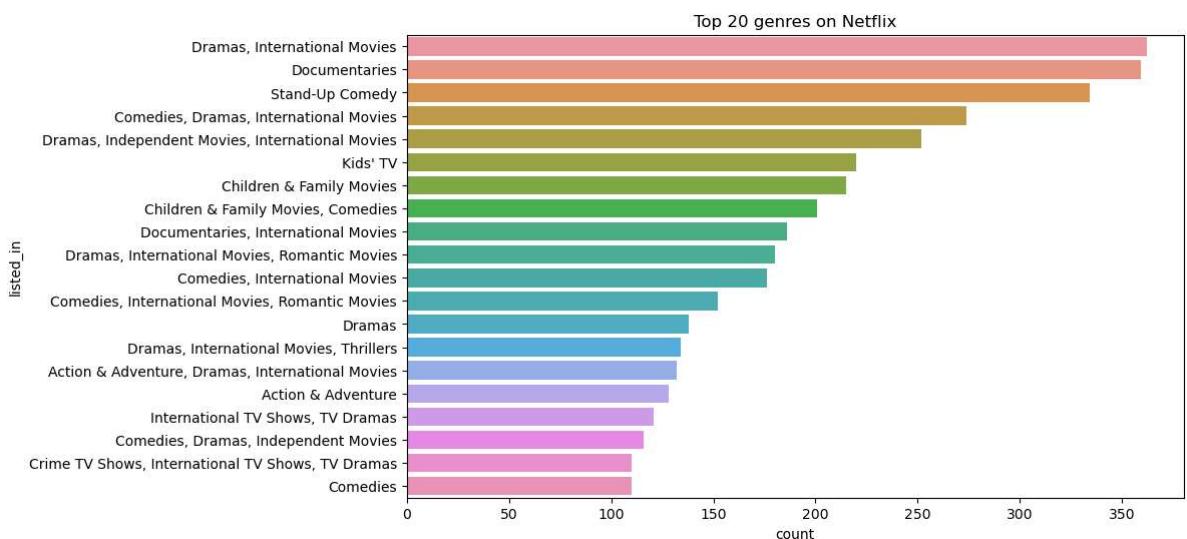
```
In [86]: plt.figure(figsize=(10,6))
sns.countplot(x='release_year',order=data['release_year'].value_counts().index[0:20])
plt.title('Content release in years on Netflix vs count')
```

```
Out[86]: Text(0.5, 1.0, 'Content release in years on Netflix vs count')
```



```
In [88]: plt.figure(figsize=(10,6))
sns.countplot(y='listed_in',order=data['listed_in'].value_counts().index[0:20],data=data)
plt.title('Top 20 genres on Netflix')
```

Out[88]: Text(0.5, 1.0, 'Top 20 genres on Netflix')



In []: