```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df=pd.read_csv("/netflix_titles.csv (1).zip")
```

```
df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
					Sami Bouajila						Crime TV	

```
df.shape # tells us the rows and columns 0f the dataset
```

```
(8807, 12)
```

```
df.describe() # tells us the some basic stats
```

	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

```
df.info() # tells us datatypes of our columns
```

```
<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
```

Missing values

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

```
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
```

Adjust Data Types

```
df['date_added'] = pd.to_datetime(df['date_added'])
```

```
df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
					Sami Bouajila						Crime TV	

Handling Missing Values

```
df.fillna({'rating':'Unavailable','cast':'Unavailable','country':'Unavailable','director':'Unavailable'},inplace=True)
df.isnull().sum()
```

```
show_id          0
type             0
title            0
director         0
cast             0
country          0
date_added      10
release_year     0
rating           0
duration         3
listed_in        0
description      0
dtype: int64
```

```
df[df.date_added.isnull()]
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descript
6066	s6067	TV Show	A Young Doctor's Notebook and Other Stories	Unavailable	Daniel Radcliffe, Jon Hamm, Adam Godley, Chris...	United Kingdom	NaT	2013	TV-MA	2 Seasons	British TV Shows, TV Comedies, TV Dramas	Set during Russ Revolution, comi
6174	s6175	TV Show	Anthony Bourdain: Parts Unknown	Unavailable	Anthony Bourdain	United States	NaT	2018	TV-PG	5 Seasons	Docuseries	This C original se has c Anthony Bo
6795	s6796	TV Show	Frasier	Unavailable	Kelsey Grammer, Jane Leeves, David Hyde Pierce...	United States	NaT	2003	TV-PG	11 Seasons	Classic & Cult TV, TV Comedies	Frasier Cr is a snooty lovable Sea
6806	s6807	TV Show	Friends	Unavailable	Jennifer Aniston, Courteney Cox, Lisa Kudrow, ...	United States	NaT	2003	TV-14	10 Seasons	Classic & Cult TV, TV Comedies	This hit sitc follows m misadventu
6901	s6902	TV Show	Gunslinger Girl	Unavailable	Yuuka Nanri, Kanako Mitsuhashi, Eri Sendai, Am...	Japan	NaT	2008	TV-14	2 Seasons	Anime Series, Crime TV Shows	On the surfa the So Welf Agency app
												A wacky ra

```
most_recent_entry_date=df['date_added'].max()
df.fillna({'date_added':most_recent_entry_date},inplace=True)
```

```
<ipython-input-19-23eda5c4d929>:2: DeprecationWarning: In a future version, `df.iloc[:, i] = newvals` will attempt to set the values
  df.fillna({'date_added':most_recent_entry_date},inplace=True)
```

Start coding or generate with AI.

Additional data cleaning

```
df[df.duration.isnull()]
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
5541	s5542	Movie	Louis C.K. 2017	Louis C.K.	Louis C.K.	United States	2017-04-04	2017	74 min	NaN	Movies	Louis C.K. muses on religion, eternal love, gi...
5794	s5795	Movie	Louis C.K.: Hilarious	Louis C.K.	Louis C.K.	United States	2016-09-16	2010	84 min	NaN	Movies	Emmy-winning comedy writer Louis C.K. brings h...
			Louis									

Start coding or generate with AI.

Check to make sure there is no other content with the same director to avoid accidental overwritring

```
df[df.director == 'Louis C.K.'].head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
5541	s5542	Movie	Louis C.K. 2017	Louis C.K.	Louis C.K.	United States	2017-04-04	2017	74 min	NaN	Movies	Louis C.K. muses on religion, eternal love, gi...
5794	s5795	Movie	Louis C.K.: Hilarious	Louis C.K.	Louis C.K.	United States	2016-09-16	2010	84 min	NaN	Movies	Emmy-winning comedy writer Louis C.K. brings h...
			Louis									

Overwrite and Check

```
df.loc[df['director']== 'Louis C.K','rating']='Unavailable'
df[df.director == 'Louis C.K'].head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description

```
#Loc helps us accessing the columns by names
df.loc[df['director']=='Louis C.K.','duration']=df['rating']
df[df.director == 'Louis C.K'].head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description

Start coding or generate with AI.

Visualizations

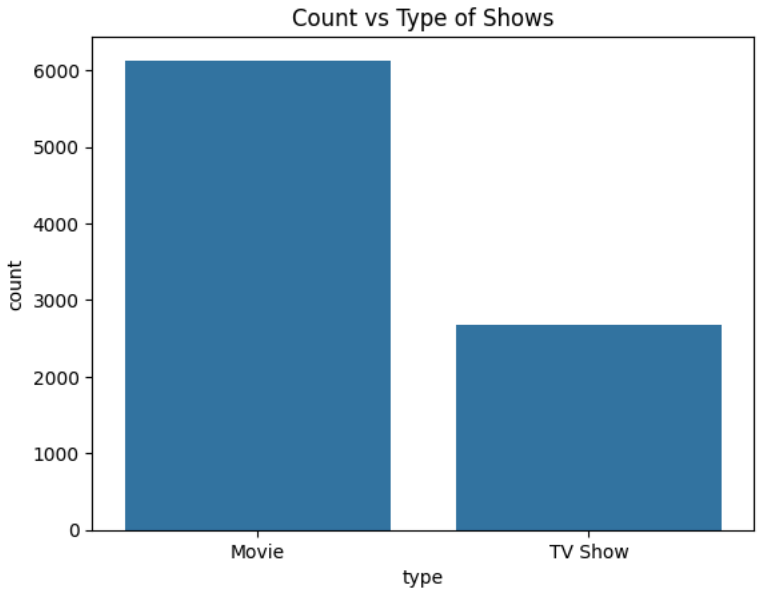
Let's take a look at types of Shows that has been watched on Netflix

```
df.type.value_counts()
```

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

```
# countplot helps us to plot counts of each category
sns.countplot(x='type',data = df)
plt.title('Count vs Type of Shows')
```

```
Text(0.5, 1.0, 'Count vs Type of Shows')
```



On Netflix there are more no. of Movies as compared to Tv shows

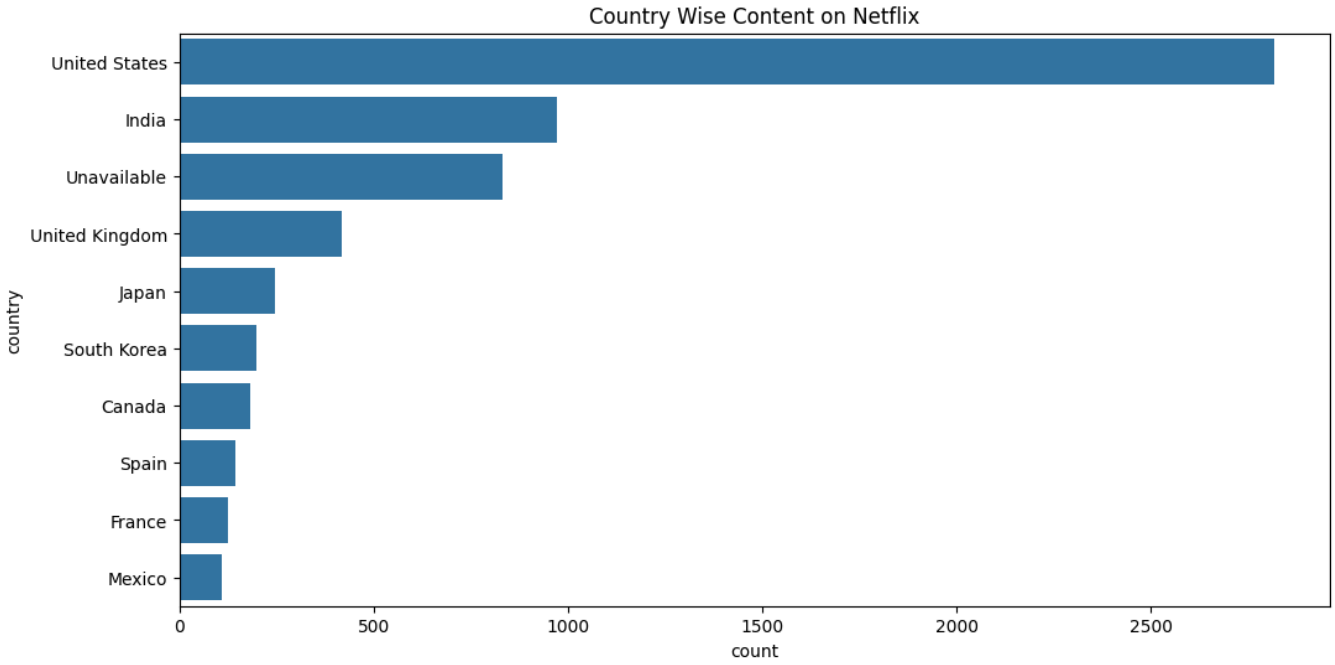
Country Analysis

```
df['country'].value_counts().head(10)
```

```
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
```

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

```
Text(0.5, 1.0, 'Country Wise Content on Netflix')
```

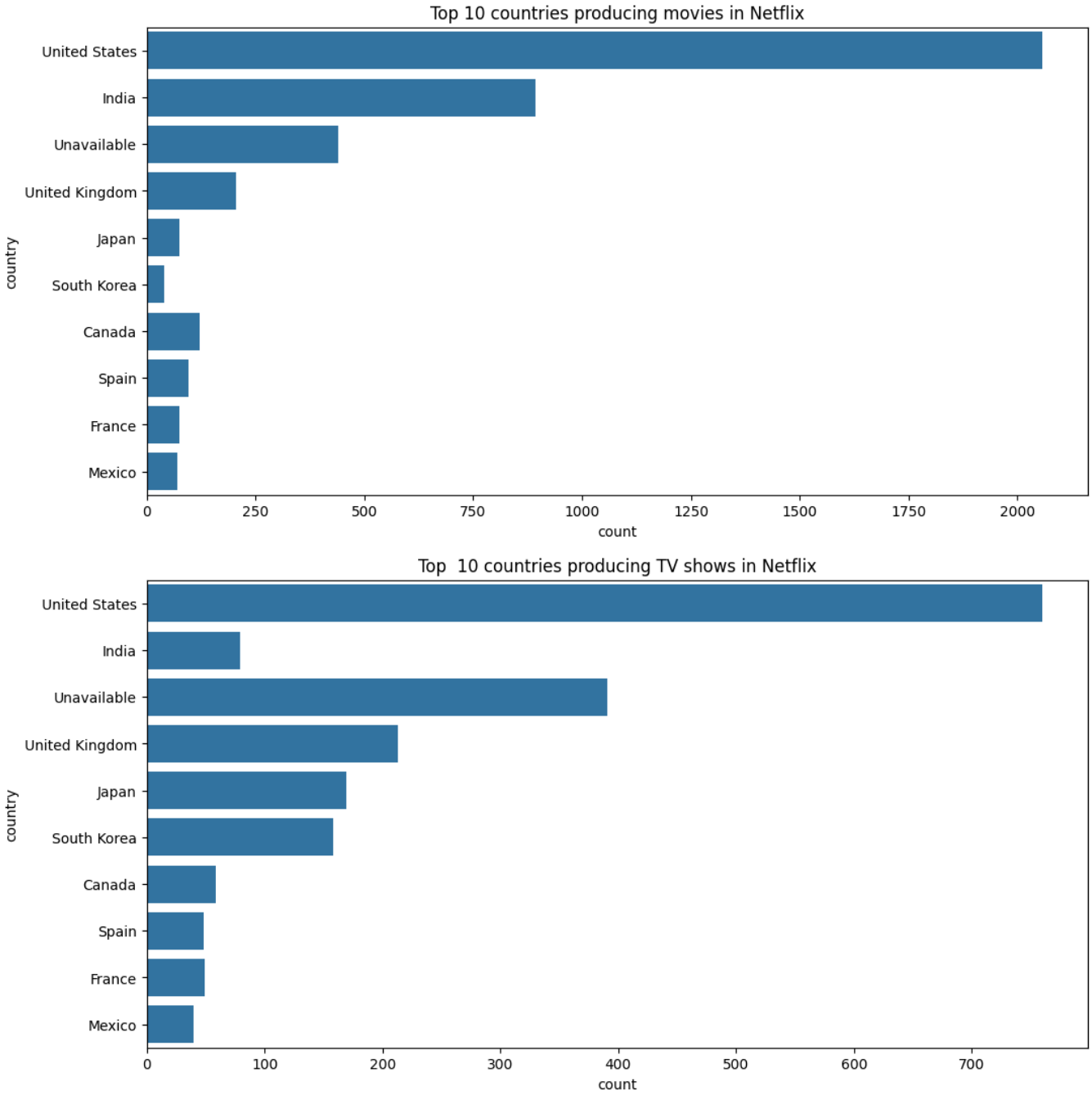



```
#now checking type of content based on country
movie_coutnries = df[df['type']=='Movie']
tv_show_countries = df[df['type']=='TV Show']
```

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

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

```
Text(0.5, 1.0, 'Top  10 countries producing TV shows in Netflix')
```



Start coding or generate with AI.

Lets Check What are the major ratings given to Netflix Shows

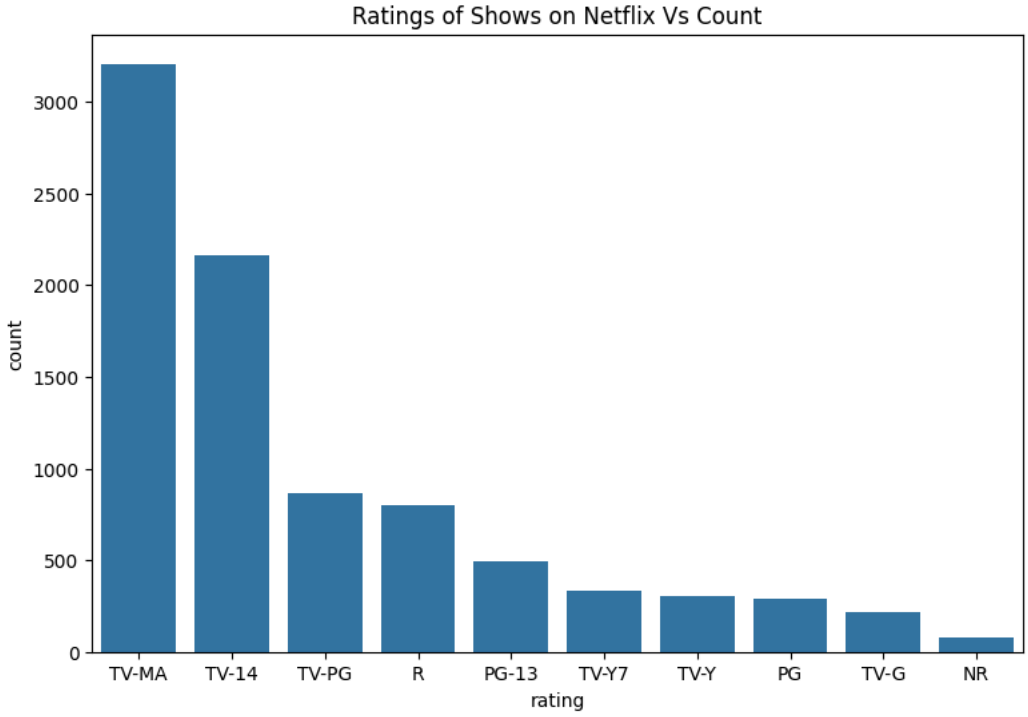
```
df.rating.value_counts()
```

```
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
```

```
plt.figure(figsize =(9,6))
sns.countplot(x='rating',order = df['rating'].value_counts().index[0:10],data = df)
plt.title('Ratings of Shows on Netflix Vs Count')
```

```
Text(0.5, 1.0, 'Ratings of Shows on Netflix Vs Count')
```



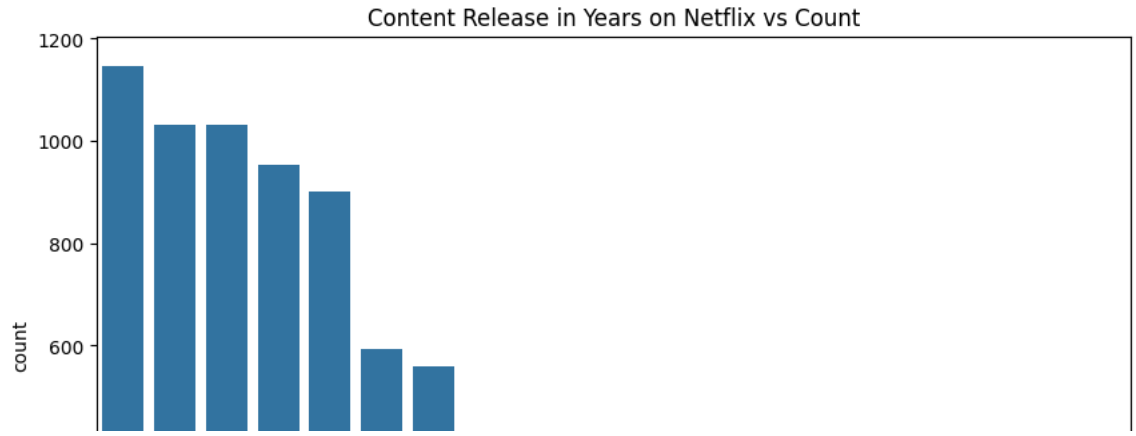
Most of the Shows has TV-MA and TV-14 ratings

```
df.release_year.value_counts()[:20]
```

```
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
```

```
plt.figure(figsize =(10,6))
sns.countplot(x='release_year',order = df['release_year'].value_counts().index[0:20],data = df)
plt.title('Content Release in Years on Netflix vs Count')
```

```
Text(0.5, 1.0, 'Content Release in Years on Netflix vs Count')
```



Popular Genres Analysis

```
plt.figure(figsize =(12,8))
sns.countplot(y='listed_in',order = df['listed_in'].value_counts().index[0:20],data = df)
plt.title('Top 20 Genres on Netflix')
```

```
Text(0.5, 1.0, 'Top 20 Genres on Netflix')
```