

```
from google.colab import drive
drive.mount('/content/drive')
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Mounted at /content/drive

```
df=pd.read_csv('/content/drive/MyDrive/dataset - netflix1.csv')
df.head()
```

| | show_id | type | title | director | country | date_added | release_year | rating | d |
|---|---------|---------|----------------------|-----------------|---------------|------------|--------------|--------|---|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | United States | 9/25/2021 | 2020 | PG-13 | |
| 1 | s3 | TV Show | Ganglands | Julien Leclercq | France | 9/24/2021 | 2021 | TV-MA | 1 |
| 2 | s6 | TV Show | Midnight Mass | Mike Flanagan | United States | 9/24/2021 | 2021 | TV-MA | 1 |

```
df.rating.unique()
```

```
array(['PG-13', 'TV-MA', 'TV-PG', 'TV-14', 'TV-Y7', 'TV-Y', 'PG', 'TV-G',
      'R', 'G', 'NC-17', 'NR', 'TV-Y7-FV', 'UR'], dtype=object)
```

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

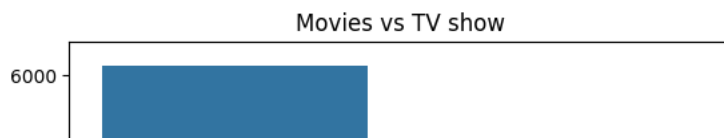
```
TV-MA      3205
TV-14      2157
TV-PG       861
R           799
PG-13       490
TV-Y7       333
TV-Y        306
PG          287
TV-G        220
NR           79
G           41
TV-Y7-FV     6
NC-17        3
UR           3
Name: rating, dtype: int64
```

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

```
Movie      6126
TV Show    2664
Name: type, dtype: int64
```

```
sns.countplot(x='type',data=df)
plt.title("Movies vs TV show")
```

```
Text(0.5, 1.0, 'Movies vs TV show')
```



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

```
United States    3240
India            1057
United Kingdom   638
Pakistan         421
Not Given        287
...
Iran             1
West Germany     1
Greece           1
Zimbabwe         1
Soviet Union     1
Name: country, Length: 86, dtype: int64
```

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

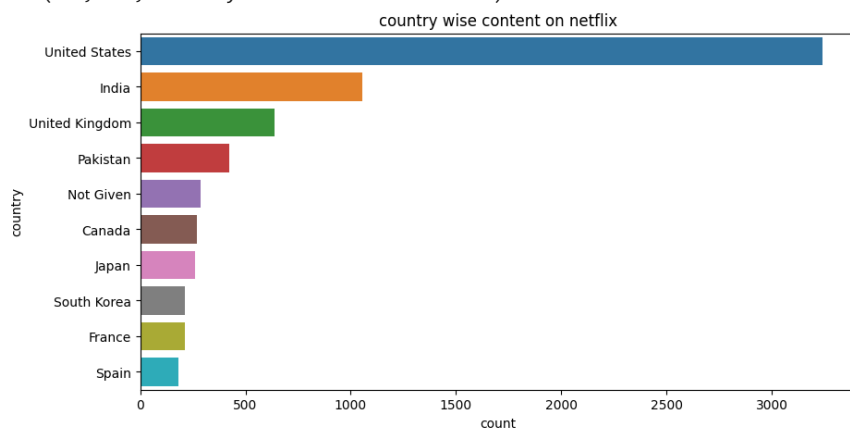
```
United States    3240
India            1057
United Kingdom   638
Pakistan         421
Not Given        287
Canada           271
Japan            259
South Korea      214
France           213
Spain            182
Name: country, dtype: int64
```

```
plt.figure(figsize=(10,5))
```

```
sns.countplot(y='country',order=df['country'].value_counts().index[0:10],data=df)
```

```
plt.title("country wise content on netflix")
```

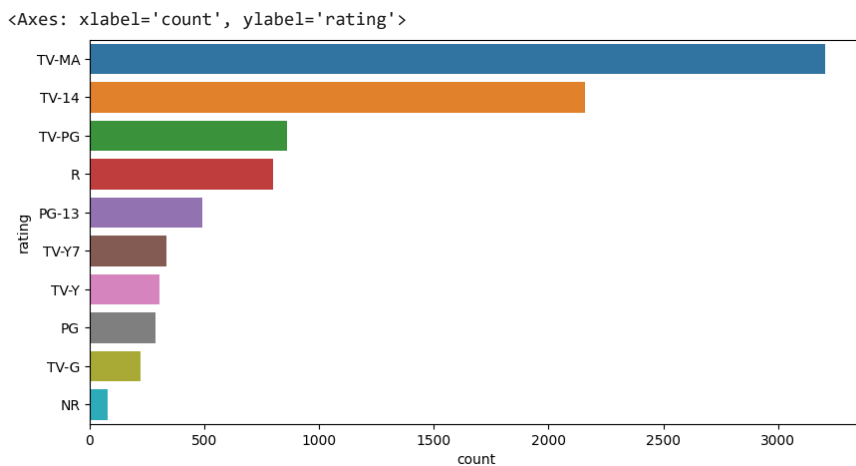
```
Text(0.5, 1.0, 'country wise content on netflix')
```



It is seen that the United States is producing most of the content on Netflix as compared to other countries. It can also be seen India is producing is second in producing the content.

```
plt.figure(figsize=(10,5))
```

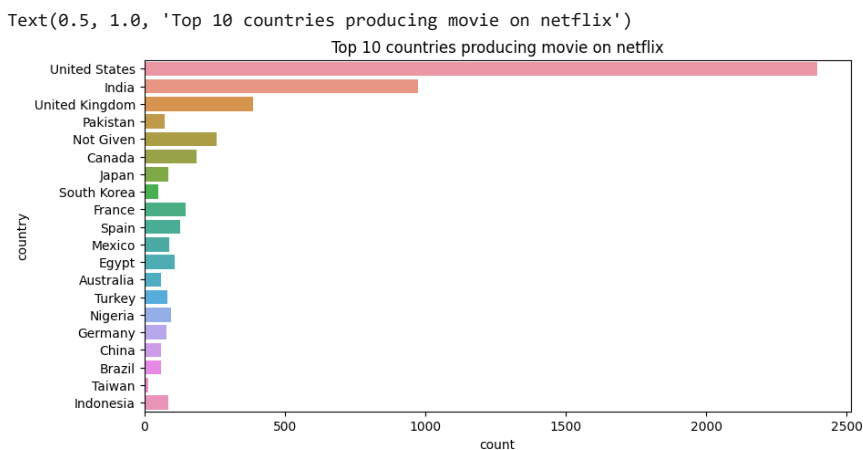
```
sns.countplot(y='rating',order=df['rating'].value_counts().index[0:10],data=df)
```



It is seen that the United States is producing most of the content on Netflix as compared to other countries. It can also be seen India is producing is second in producing the content.

```
movies_country=df[df['type']=='Movie']
tv_show_country=df[df['type']=='TV Show']
```

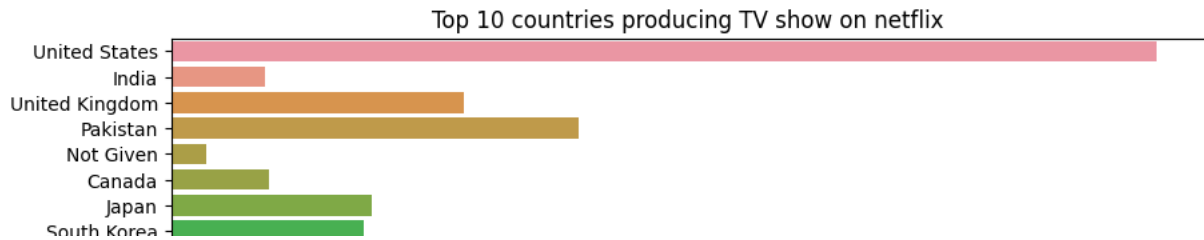
```
plt.figure(figsize=(10,5))
sns.countplot(y='country',order=df['country'].value_counts().index[:20],data=movies_country)
plt.title("Top 10 countries producing movie on netflix")
```



United states producing Movies most on the netflix compared to another country.

```
plt.figure(figsize=(10,5))
sns.countplot(y='country',order=df['country'].value_counts().index[:20],data=tv_show_country)
plt.title("Top 10 countries producing TV show on netflix")
```

```
Text(0.5, 1.0, 'Top 10 countries producing TV show on netflix')
```

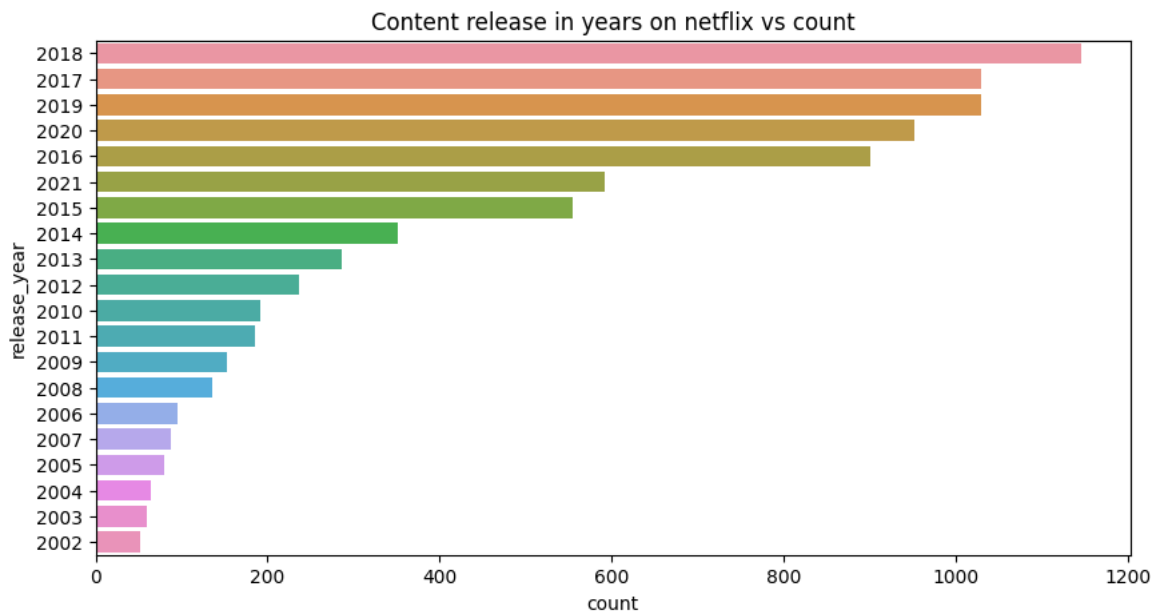


Double-click (or enter) to edit

```
plt.figure(figsize=(10,5))
```

```
sns.countplot(y='release_year',order=df['release_year'].value_counts().index[: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')
```



It is seen that Netflix is releasing increasing their content as years are passing by.

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

```
2018    1146
2017    1030
2019    1030
2020     953
2016     901
2021     592
2015     555
2014     352
2013     286
2012     236
2010     192
2011     185
2009     152
2008     135
2006      96
2007      88
2005      80
2004      64
2003      59
2002      51
Name: release_year, dtype: int64
```

```
plt.figure(figsize=(10,5))
sns.countplot(y='listed_in',order=df['listed_in'].value_counts().index[0:20],data=df)
plt.title('Top 20 genre on Netflix')
```

Text(0.5, 1.0, 'Top 20 genre on Netflix')

