In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
from IPython import get_ipython
import warnings
warnings.filterwarnings("ignore")
```

In [2]:

```
data = pd.read_csv('netflix_titles.csv')
```

In [3]:

```
data.head()
```

Out[3]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	d
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	S
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	S

In [4]:

```
1 data.tail()
```

Out[4]:

	show_id	type	title	director	cast	country	date_added	release_year	rati
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	I
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-

In [5]:

```
1 data.shape
```

Out[5]:

```
(8807, 12)
```

In [6]:

```
1 data.columns
```

Out[6]:

```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_ad
ded',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object')
```

In [7]:

```
1 data.duplicated().sum()
```

Out[7]:

```
0
```

In [8]:

```
1 data.isnull().sum()
```

Out[8]:

```
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 [9]:

```
1 data.director.fillna(value="NA",inplace=True)
```

In [10]:

```
1 data.cast.fillna(value="NA",inplace=True)
```

In [11]:

```
1 data.country.fillna(value="NA",inplace=True)
```

In [12]:

```
1 data.date_added.fillna(value="NA",inplace=True)
```

In [13]:

```
1 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      8807 non-null   object
 4   cast          8807 non-null   object
 5   country       8807 non-null   object
 6   date_added    8807 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 [14]:

```
1 data.describe()
```

Out[14]:

	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 [15]:

```
1 data.nunique()
```

Out[15]:

```
show_id         8807
type               2
title           8804
director        4529
cast            7693
country          749
date_added      1768
release_year      74
rating            17
duration         220
listed_in        514
description     8775
dtype: int64
```

In [16]:

```
1 data.type.unique()
```

Out[16]:

```
array(['Movie', 'TV Show'], dtype=object)
```

In [17]:

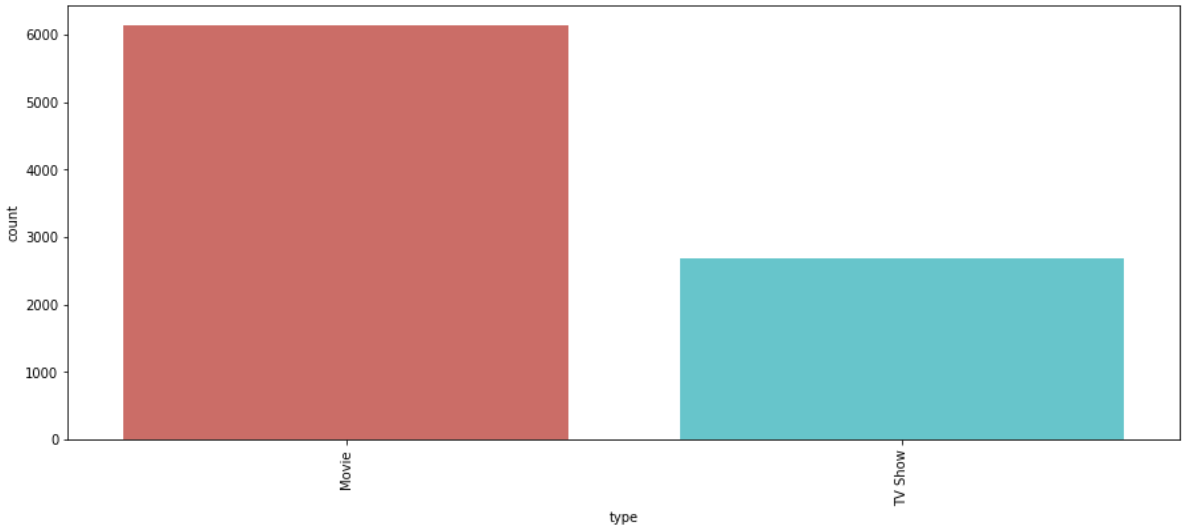
```
1 data.type.value_counts()
```

Out[17]:

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

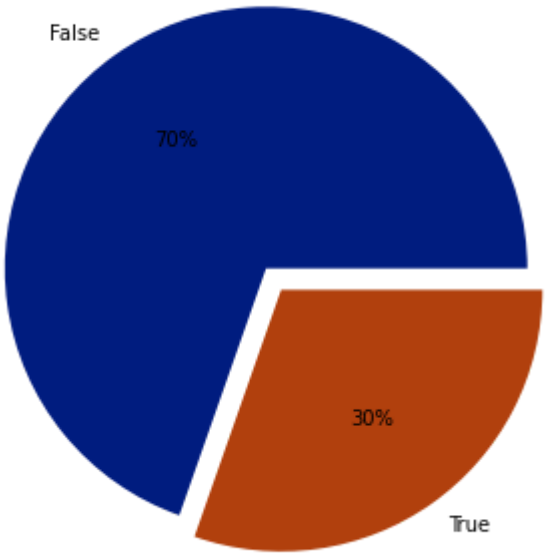
In [18]:

```
plt.figure(figsize=(15,6))
sns.countplot('type', data = data, palette = 'hls')
plt.xticks(rotation = 90)
plt.show()
```



In [19]:

```
plt.figure(figsize=(15,6))
palette_color = sns.color_palette('dark')
explode = [0, 0.1]
plt.pie(data['type'].value_counts(), labels= ['False', 'True'],
        colors=palette_color, explode=explode,
        autopct='%.0f%%')
plt.show()
```



In [20]:

```
last_years = data[["type","release_year"]]
```

In [21]:

```
last_years=last_years[last_years["release_year"] >= 2010]
```

In [22]:

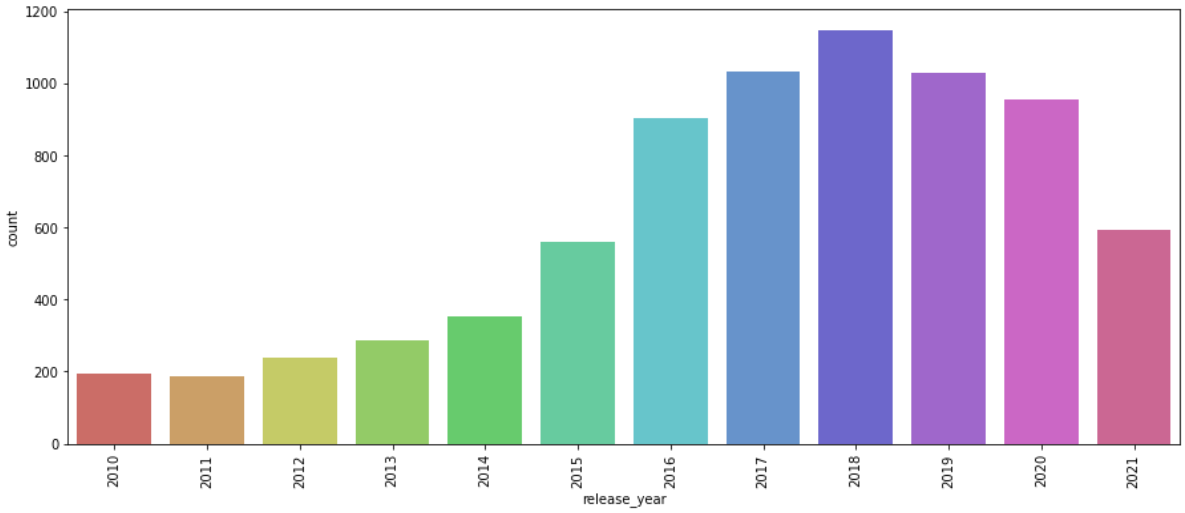
```
last_years.release_year.value_counts().sort_index()
```

Out[22]:

```
2010     194
2011     185
2012     237
2013     288
2014     352
2015     560
2016     902
2017    1032
2018    1147
2019    1030
2020     953
2021     592
Name: release_year, dtype: int64
```

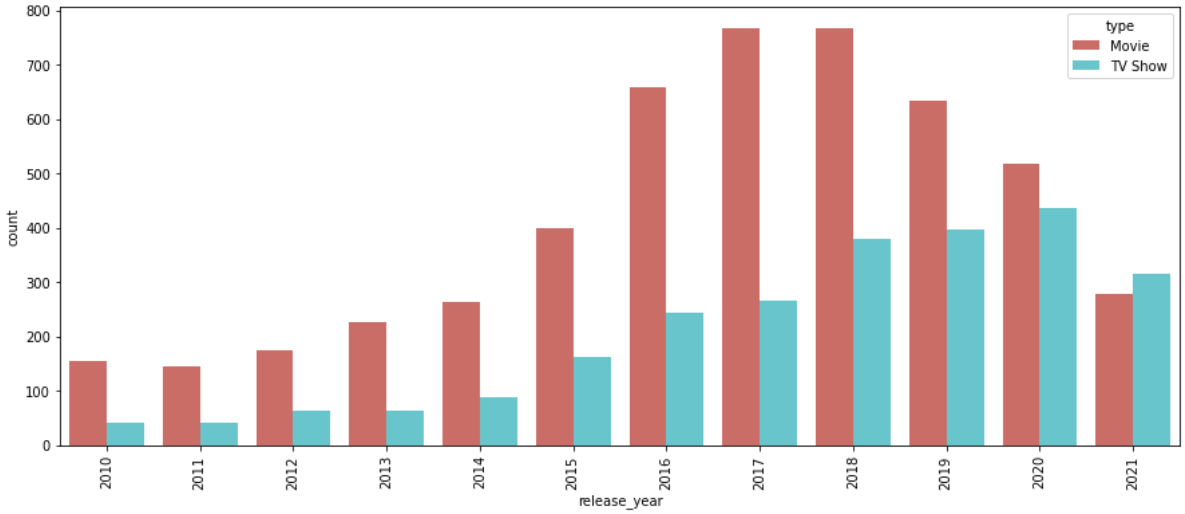
In [23]:

```
plt.figure(figsize=(15,6))
sns.countplot('release_year', data = last_years,
              palette = 'hls')
plt.xticks(rotation = 90)
plt.show()
```



In [24]:

```
plt.figure(figsize=(15,6))
sns.countplot('release_year', data = last_years,
              hue = 'type' , palette = 'hls')
plt.xticks(rotation = 90)
plt.show()
```



In [25]:

```
data.rating.unique()
```

Out[25]:

```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
       'TV-G', 'G', 'NC-17', '74 min', '84 min', '66 min', 'NR', nan,
       'TV-Y7-FV', 'UR'], dtype=object)
```

In [26]:

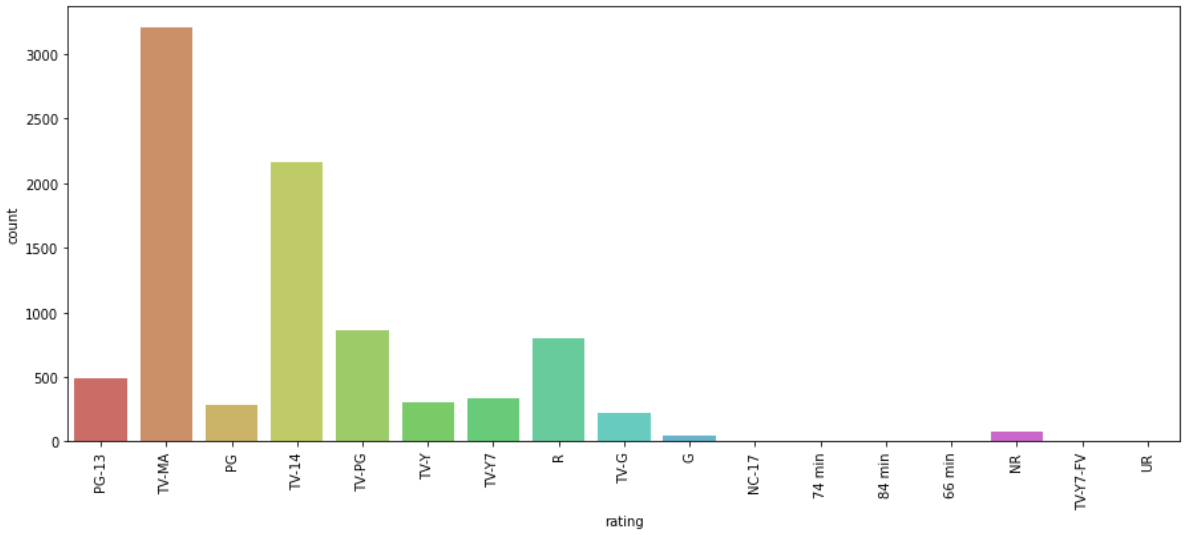
```
1 data.rating.value_counts()
```

Out[26]:

```
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
NC-17          3
UR             3
74 min         1
84 min         1
66 min         1
Name: rating, dtype: int64
```

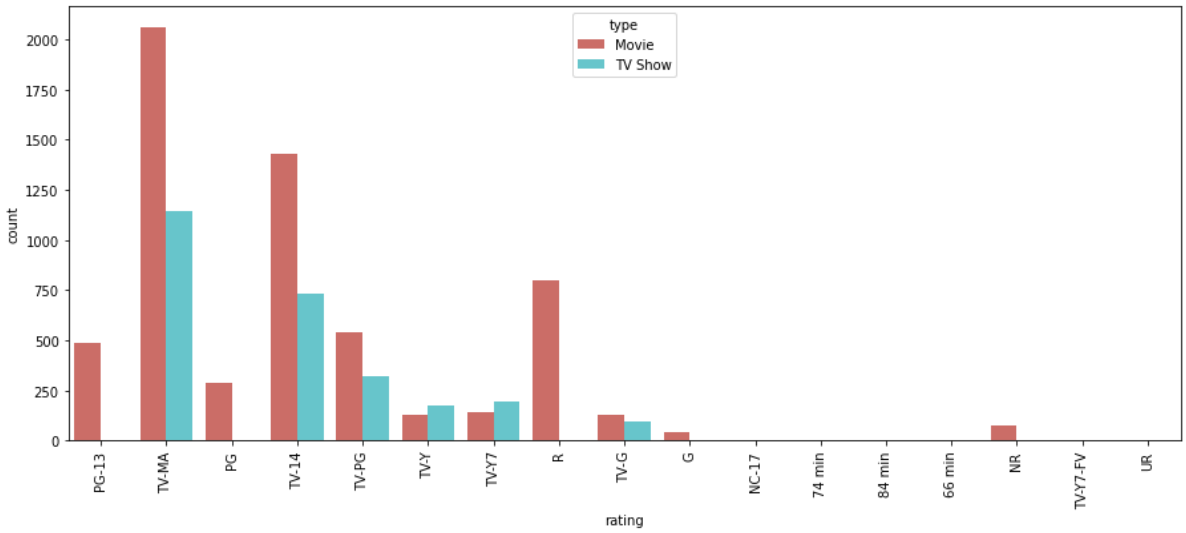
In [27]:

```
1 plt.figure(figsize=(15,6))
2 sns.countplot('rating', data = data, palette = 'hls')
3 plt.xticks(rotation = 90)
4 plt.show()
```



In [28]:

```
plt.figure(figsize=(15,6))
sns.countplot('rating', data = data, hue = 'type' ,
              palette = 'hls')
plt.xticks(rotation = 90)
plt.show()
```



In [29]:

```
new_categories = {
    'TV-PG': 'Parental Guidance',
    'TV-MA': 'Mature Audience',
    'TV-Y7-FV': 'Teens',
    'TV-Y7': 'Teens',
    'TV-14': 'Teens',
    'R': 'Mature Audience',
    'TV-Y': 'General Audience',
    'NR': 'Mature Audience',
    'PG-13': 'Teens',
    'TV-G': 'General Audience',
    'PG': 'Teens',
    'G': 'General Audience',
    'UR': 'Mature Audience',
    'NC-17': 'Mature Audience'
}
data["rating"] = data['rating'].replace(new_categories)
```

In [30]:

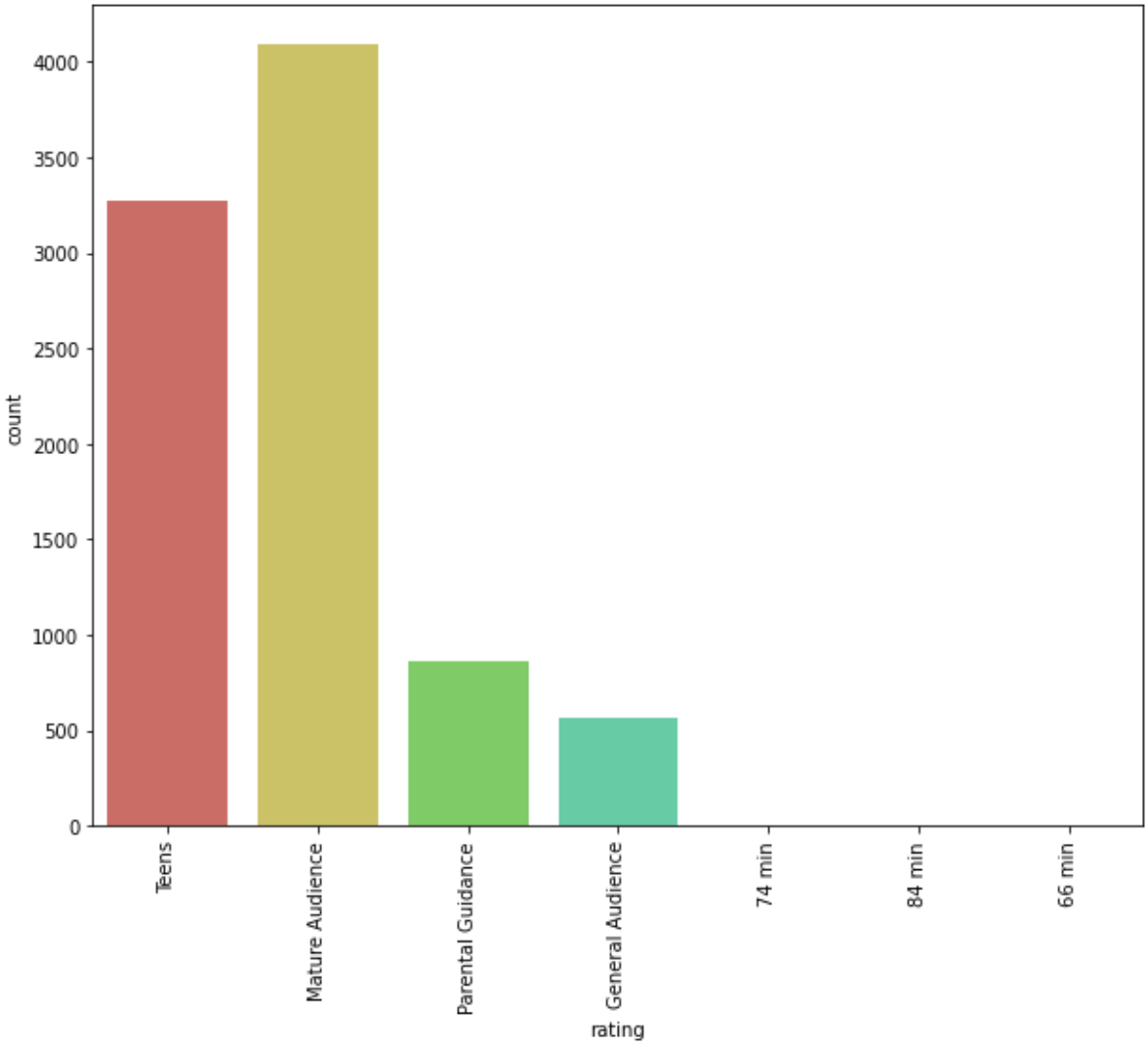
```
1 data.head()
```

Out[30]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NA	United States	September 25, 2021	2020	Teens
1	s2	TV Show	Blood & Water	NA	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	Mature Audience
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NA	September 24, 2021	2021	Mature Audience
3	s4	TV Show	Jailbirds New Orleans	NA	NA	NA	September 24, 2021	2021	Mature Audience
4	s5	TV Show	Kota Factory	NA	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	Mature Audience

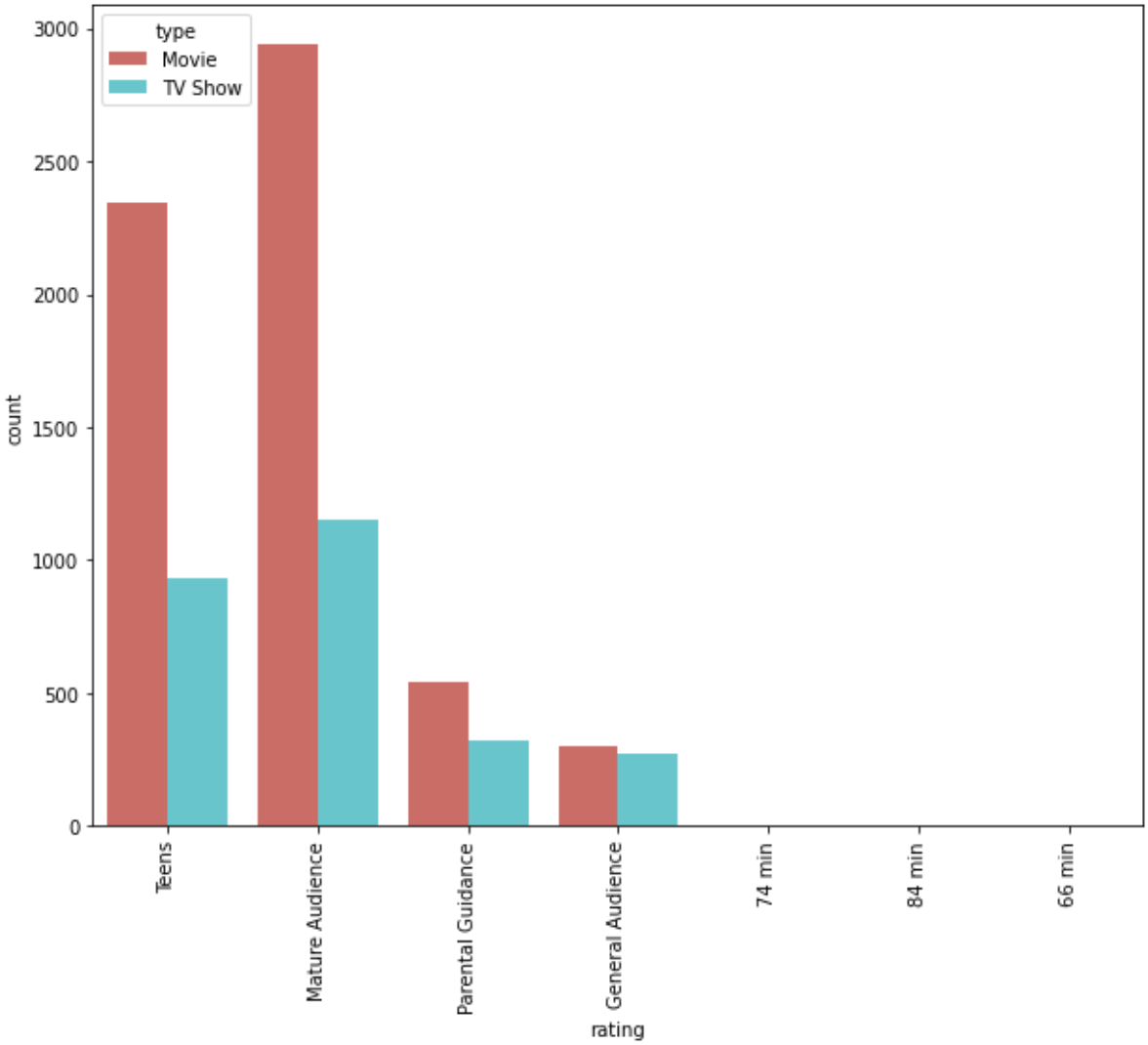
In [31]:

```
plt.figure(figsize=(10,8))
sns.countplot("rating", data=data, palette="hls")
plt.xticks(rotation = 90)
plt.show()
```



In [32]:

```
plt.figure(figsize=(10,8))
sns.countplot("rating", data=data, hue =  'type' ,
              palette="hls")
plt.xticks(rotation = 90)
plt.show()
```



In [33]:

```
1 data.country.unique()
```

Out[33]:

```
array(['United States', 'South Africa', 'NA', 'India',
       'United States, Ghana, Burkina Faso, United Kingdom, Germany, Eth
iopia',
       'United Kingdom', 'Germany, Czech Republic', 'Mexico', 'Turkey',
       'Australia', 'United States, India, France', 'Finland',
       'China, Canada, United States',
       'South Africa, United States, Japan', 'Nigeria', 'Japan',
       'Spain, United States', 'France', 'Belgium',
       'United Kingdom, United States', 'United States, United Kingdom',
       'France, United States', 'South Korea', 'Spain',
       'United States, Singapore', 'United Kingdom, Australia, France',
       'United Kingdom, Australia, France, United States',
       'United States, Canada', 'Germany, United States',
       'South Africa, United States', 'United States, Mexico',
       'United States, Italy, France, Japan',
       'United States, Italy, Romania, United Kingdom',
       'Australia, United States', 'Argentina, Venezuela',
```

In [34]:

```
1 data.country.value_counts()
```

Out[34]:

```
United States                                2818
India                                         972
NA                                            831
United Kingdom                                419
Japan                                         245
                                             ...
Romania, Bulgaria, Hungary                      1
Uruguay, Guatemala                              1
France, Senegal, Belgium                        1
Mexico, United States, Spain, Colombia          1
United Arab Emirates, Jordan                    1
Name: country, Length: 749, dtype: int64
```

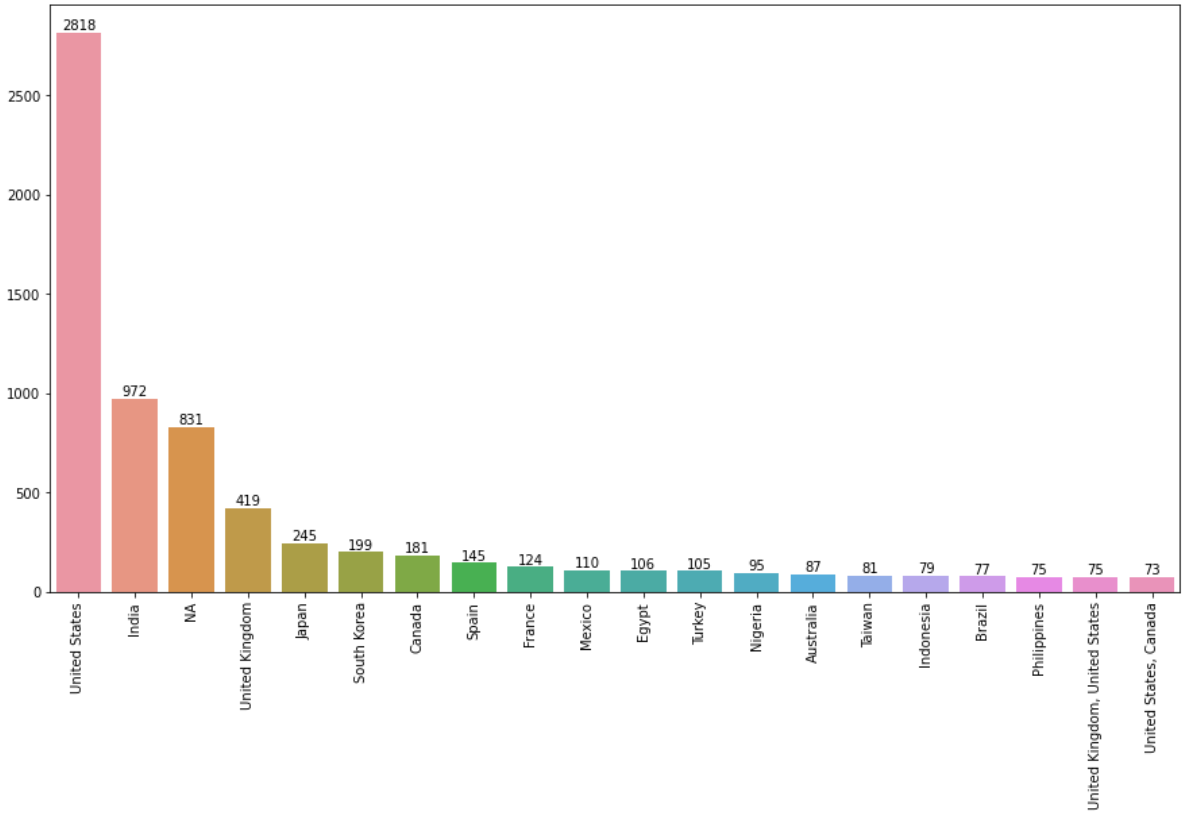
In [35]:

```
plt.figure(figsize=(15,8))

ax = sns.barplot(x=data.country.value_counts()[:20].index ,
                 y=data.country.value_counts()[:20].values )

ax.set_xticklabels(data.country.value_counts()[:20].index,
                   rotation=90)

for i in ax.containers:
    ax.bar_label(i);
```



In [37]:

```
plt.figure(figsize=(10,8))
sns.countplot("country", data=data.head(100), hue =  'type' ,
              palette="hls")
plt.xticks(rotation = 90)
plt.show()
```