In [1]:

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

In [2]:

```python
data = pd.read_csv(r"C:\Users\akash.bana\Desktop\Akash_backup\Akash\Scaler\Python librar
data
```

Out[2]:

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

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

In [3]:

8807 rows × 12 columns
```
data.head()
```

Out[3]:

	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
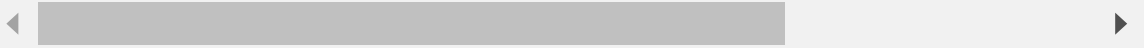
In [4]:

```
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   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 [6]:

```
data.shape
```

Out[6]:

```
(8807, 12)
```

In [7]:

```
data.describe()
```

Out[7]:

	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 [8]:

```python
data.describe(include="object")
```

Out[8]:

	show_id	type	title	director	cast	country	date_added	rating	durati
count	8807	8807	8807	6173	7982	7976	8797	8803	88
unique	8807	2	8807	4528	7692	748	1767	17	2
top	s1	Movie	Dick Johnson Is Dead	Rajiv Chilaka	David Attenborough	United States	January 1, 2020	TV-MA	Seas
freq	1	6131	1	19	19	2818	109	3207	17

Missing values

In [222]:

```python
round(data.isna().sum()/len(data['title'])*100,2)
```

Out[222]:

```
show_id         0.00
type            0.00
title           0.00
director       29.91
cast            9.37
country         9.44
date_added      0.11
release_year    0.00
rating          0.05
duration        0.03
listed_in       0.00
description     0.00
dtype: float64
```

In [224]:

```python
data['director'].nunique()
```

Out[224]:

```
4528
```

In [14]:

```python
data["type"].value_counts()
```

Out[14]:

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

Splitting cast

In [5]:

```python
cast = data["cast"].apply(lambda x: str(x).split(',')).tolist()
df1 = pd.DataFrame(cast, index=data['title'])
df1 = df1.stack()
df1 = pd.DataFrame(df1)
df1 = df1.reset_index()
df1 = df1[['title',0]]
df1 = df1.rename(columns={'title':'Title',0:'cast'})
df1['cast'] = df1['cast'].replace({'nan':np.nan})
df1
```

Out[5]:

	Title	cast
0	Dick Johnson Is Dead	NaN
1	Blood & Water	Ama Qamata
2	Blood & Water	Khosi Ngema
3	Blood & Water	Gail Mabalane
4	Blood & Water	Thabang Molaba
...	...	...
64946	Zubaan	Manish Chaudhary
64947	Zubaan	Meghna Malik
64948	Zubaan	Malkeet Rauni
64949	Zubaan	Anita Shabdish
64950	Zubaan	Chittaranjan Tripathy

64951 rows × 2 columns

Splitting genre

In [6]:

```python
genre = data["listed_in"].apply(lambda x: str(x).split(',')).tolist()
df2 = pd.DataFrame(genre, index=data['title'])
df2 = pd.DataFrame(df2.stack())
df2 = df2.reset_index()
df2 = df2[["title",0]]
df2 = df2.rename(columns={'title':'Title',0:'genre'})
df2['genre'] = df2['genre'].replace({'nan':np.nan})
df2
```

Out[6]:

	Title	genre
0	Dick Johnson Is Dead	Documentaries
1	Blood & Water	International TV Shows
2	Blood & Water	TV Dramas
3	Blood & Water	TV Mysteries
4	Ganglands	Crime TV Shows
...	...	...
19318	Zoom	Children & Family Movies
19319	Zoom	Comedies
19320	Zubaan	Dramas
19321	Zubaan	International Movies
19322	Zubaan	Music & Musicals

19323 rows × 2 columns

Splitting director

In [7]:

```python
directors = data['director'].apply(lambda x: str(x).split(',')).tolist()
df3 = pd.DataFrame(directors, index=data['title'])
df3 = pd.DataFrame(df3.stack())
df3 = df3.reset_index()
df3 = df3[['title',0]]
df3 = df3.rename(columns={'title':'Title',0:"directors"})
df3['directors'] = df3['directors'].replace({'nan':np.nan})
df3
```

Out[7]:

	Title	directors
0	Dick Johnson Is Dead	Kirsten Johnson
1	Blood & Water	NaN
2	Ganglands	Julien Leclercq
3	Jailbirds New Orleans	NaN
4	Kota Factory	NaN
...	...	...
9607	Zodiac	David Fincher
9608	Zombie Dumb	NaN
9609	Zombieland	Ruben Fleischer
9610	Zoom	Peter Hewitt
9611	Zubaan	Mozez Singh

9612 rows × 2 columns

Splitting country

In [68]:

```python
countries = data['country'].apply(lambda x: str(x).split(',')).tolist()
df4 = pd.DataFrame(countries, index=data['title'])
df4 = pd.DataFrame(df4.stack())
df4 = df4.reset_index()
df4 = df4[['title',0]]
df4 = df4.rename(columns={'title':'Title',0:'country'})
df4['country'] = df4['country'].replace({'nan':np.nan})
df4
```

Out[68]:

	Title	country
0	Dick Johnson Is Dead	United States
1	Blood & Water	South Africa
2	Ganglands	NaN
3	Jailbirds New Orleans	NaN
4	Kota Factory	India
...	...	...
10845	Zodiac	United States
10846	Zombie Dumb	NaN
10847	Zombieland	United States
10848	Zoom	United States
10849	Zubaan	India

10850 rows × 2 columns

Merge all

In [69]:

```
df12 = df1.merge(df2,on='Title')
df12
```

Out[69]:

	Title	cast	genre
0	Dick Johnson Is Dead	NaN	Documentaries
1	Blood & Water	Ama Qamata	International TV Shows
2	Blood & Water	Ama Qamata	TV Dramas
3	Blood & Water	Ama Qamata	TV Mysteries
4	Blood & Water	Khosi Ngema	International TV Shows
...	...	...	...
149507	Zubaan	Anita Shabdish	International Movies
149508	Zubaan	Anita Shabdish	Music & Musicals
149509	Zubaan	Chittaranjan Tripathy	Dramas
149510	Zubaan	Chittaranjan Tripathy	International Movies
149511	Zubaan	Chittaranjan Tripathy	Music & Musicals

149512 rows × 3 columns

In [70]:

```
df123 = df12.merge(df3, on = 'Title')
df123
```

Out[70]:

	Title	cast	genre	directors
0	Dick Johnson Is Dead	NaN	Documentaries	Kirsten Johnson
1	Blood & Water	Ama Qamata	International TV Shows	NaN
2	Blood & Water	Ama Qamata	TV Dramas	NaN
3	Blood & Water	Ama Qamata	TV Mysteries	NaN
4	Blood & Water	Khosi Ngema	International TV Shows	NaN
...	...	...	...	...
161211	Zubaan	Anita Shabdish	International Movies	Mozez Singh
161212	Zubaan	Anita Shabdish	Music & Musicals	Mozez Singh
161213	Zubaan	Chittaranjan Tripathy	Dramas	Mozez Singh
161214	Zubaan	Chittaranjan Tripathy	International Movies	Mozez Singh
161215	Zubaan	Chittaranjan Tripathy	Music & Musicals	Mozez Singh

161216 rows × 4 columns

In [71]:

```
df1234 = df123.merge(df4,on='Title')
df1234
```

Out[71]:

	Title	cast	genre	directors	country
0	Dick Johnson Is Dead	NaN	Documentaries	Kirsten Johnson	United States
1	Blood & Water	Ama Qamata	International TV Shows	NaN	South Africa
2	Blood & Water	Ama Qamata	TV Dramas	NaN	South Africa
3	Blood & Water	Ama Qamata	TV Mysteries	NaN	South Africa
4	Blood & Water	Khosi Ngema	International TV Shows	NaN	South Africa
...	...	...	...	...	...
202060	Zubaan	Anita Shabdish	International Movies	Mozez Singh	India
202061	Zubaan	Anita Shabdish	Music & Musicals	Mozez Singh	India
202062	Zubaan	Chittaranjan Tripathy	Dramas	Mozez Singh	India
202063	Zubaan	Chittaranjan Tripathy	International Movies	Mozez Singh	India
202064	Zubaan	Chittaranjan Tripathy	Music & Musicals	Mozez Singh	India

202065 rows × 5 columns

In [72]:

```python
df_rem = data[['show_id','type','title','date_added','release_year','rating','duration']
df_rem
```

Out[72]:

	show_id	type	title	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	September 25, 2021	2020	PG-13	90 min
1	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons
2	s3	TV Show	Ganglands	September 24, 2021	2021	TV-MA	1 Season
3	s4	TV Show	Jailbirds New Orleans	September 24, 2021	2021	TV-MA	1 Season
4	s5	TV Show	Kota Factory	September 24, 2021	2021	TV-MA	2 Seasons
...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	November 20, 2019	2007	R	158 min
8803	s8804	TV Show	Zombie Dumb	July 1, 2019	2018	TV-Y7	2 Seasons
8804	s8805	Movie	Zombieland	November 1, 2019	2009	R	88 min
8805	s8806	Movie	Zoom	January 11, 2020	2006	PG	88 min
8806	s8807	Movie	Zubaan	March 2, 2019	2015	TV-14	111 min

8807 rows × 7 columns

In [74]:

```python
df = df1234.merge(df_rem,left_on='Title',right_on='title')
df
```

Out[74]:

	Title	cast	genre	directors	country	show_id	type	title	d
0	Dick Johnson Is Dead	NaN	Documentaries	Kirsten Johnson	United States	s1	Movie	Dick Johnson Is Dead	
1	Blood & Water	Ama Qamata	International TV Shows	NaN	South Africa	s2	TV Show	Blood & Water	
2	Blood & Water	Ama Qamata	TV Dramas	NaN	South Africa	s2	TV Show	Blood & Water	
3	Blood & Water	Ama Qamata	TV Mysteries	NaN	South Africa	s2	TV Show	Blood & Water	
4	Blood & Water	Khosi Ngema	International TV Shows	NaN	South Africa	s2	TV Show	Blood & Water	
...	...	...	...	...	...	...	...	...	
202060	Zubaan	Anita Shabdish	International Movies	Mozez Singh	India	s8807	Movie	Zubaan	
202061	Zubaan	Anita Shabdish	Music & Musicals	Mozez Singh	India	s8807	Movie	Zubaan	
202062	Zubaan	Chittaranjan Tripathy	Dramas	Mozez Singh	India	s8807	Movie	Zubaan	
202063	Zubaan	Chittaranjan Tripathy	International Movies	Mozez Singh	India	s8807	Movie	Zubaan	
202064	Zubaan	Chittaranjan Tripathy	Music & Musicals	Mozez Singh	India	s8807	Movie	Zubaan	

202065 rows × 12 columns

In [75]:

```
df = df.drop(columns=['title'])
df
```

Out[75]:

	Title	cast	genre	directors	country	show_id	type	date_added
0	Dick Johnson Is Dead	NaN	Documentaries	Kirsten Johnson	United States	s1	Movie	September 25, 2021
1	Blood & Water	Ama Qamata	International TV Shows	NaN	South Africa	s2	TV Show	September 24, 2021
2	Blood & Water	Ama Qamata	TV Dramas	NaN	South Africa	s2	TV Show	September 24, 2021
3	Blood & Water	Ama Qamata	TV Mysteries	NaN	South Africa	s2	TV Show	September 24, 2021
4	Blood & Water	Khosi Ngema	International TV Shows	NaN	South Africa	s2	TV Show	September 24, 2021
...	...	...	...	...	...	...	...	..
202060	Zubaan	Anita Shabdish	International Movies	Mozez Singh	India	s8807	Movie	March 2 2019
202061	Zubaan	Anita Shabdish	Music & Musicals	Mozez Singh	India	s8807	Movie	March 2 2019
202062	Zubaan	Chittaranjan Tripathy	Dramas	Mozez Singh	India	s8807	Movie	March 2 2019
202063	Zubaan	Chittaranjan Tripathy	International Movies	Mozez Singh	India	s8807	Movie	March 2 2019
202064	Zubaan	Chittaranjan Tripathy	Music & Musicals	Mozez Singh	India	s8807	Movie	March 2 2019

202065 rows × 11 columns

Missing values

In [76]:

```
df.isna().sum()
```

Out[76]:

```
Title                0
cast              2149
genre                0
directors        50643
country          11897
show_id              0
type                 0
date_added         158
release_year         0
rating              67
duration             3
dtype: int64
```

In [77]:

```
x = df[['Title','directors','country']]
x = x.drop_duplicates(keep='first')
x
```

Out[77]:

	Title	directors	country
0	Dick Johnson Is Dead	Kirsten Johnson	United States
1	Blood & Water	NaN	South Africa
58	Ganglands	Julien Leclercq	NaN
85	Jailbirds New Orleans	NaN	NaN
87	Kota Factory	NaN	India
...	...	...	...
201976	Zodiac	David Fincher	United States
202006	Zombie Dumb	NaN	NaN
202009	Zombieland	Ruben Fleischer	United States
202023	Zoom	Peter Hewitt	United States
202041	Zubaan	Mozez Singh	India

11917 rows × 3 columns

In [90]:

```
round(x['directors'].isna().sum()/x.shape[0]*100,2)
```

Out[90]:

```
24.92
```

In [92]:

```
df['directors'] = df['directors'].fillna('unknown')
df['cast'] = df['cast'].fillna('unknown')
df['country'] = df['country'].fillna('unknown')
df
```

Out[92]:

	Title	cast	genre	directors	country	show_id	type	date_added
0	Dick Johnson Is Dead	unknown	Documentaries	Kirsten Johnson	United States	s1	Movie	September 25, 2021
1	Blood & Water	Ama Qamata	International TV Shows	unknown	South Africa	s2	TV Show	September 24, 2021
2	Blood & Water	Ama Qamata	TV Dramas	unknown	South Africa	s2	TV Show	September 24, 2021
3	Blood & Water	Ama Qamata	TV Mysteries	unknown	South Africa	s2	TV Show	September 24, 2021
4	Blood & Water	Khosi Ngema	International TV Shows	unknown	South Africa	s2	TV Show	September 24, 2021
...	...	...	...	...	...	...	...	..
202060	Zubaan	Anita Shabdish	International Movies	Mozez Singh	India	s8807	Movie	March 2 2019
202061	Zubaan	Anita Shabdish	Music & Musicals	Mozez Singh	India	s8807	Movie	March 2 2019
202062	Zubaan	Chittaranjan Tripathy	Dramas	Mozez Singh	India	s8807	Movie	March 2 2019
202063	Zubaan	Chittaranjan Tripathy	International Movies	Mozez Singh	India	s8807	Movie	March 2 2019
202064	Zubaan	Chittaranjan Tripathy	Music & Musicals	Mozez Singh	India	s8807	Movie	March 2 2019

202065 rows × 11 columns

Final dataframe

In [133]:

```
df
```

Out[133]:

	Title	cast	genre	directors	country	show_id	type	date_added
0	Dick Johnson Is Dead	unknown	Documentaries	Kirsten Johnson	United States	s1	Movie	September 25, 2021
1	Blood & Water	Ama Qamata	International TV Shows	unknown	South Africa	s2	TV Show	September 24, 2021
2	Blood & Water	Ama Qamata	TV Dramas	unknown	South Africa	s2	TV Show	September 24, 2021
3	Blood & Water	Ama Qamata	TV Mysteries	unknown	South Africa	s2	TV Show	September 24, 2021
4	Blood & Water	Khosi Ngema	International TV Shows	unknown	South Africa	s2	TV Show	September 24, 2021
...	...	...	...	...	...	...	...	..
202060	Zubaan	Anita Shabdish	International Movies	Mozez Singh	India	s8807	Movie	March 2 2019
202061	Zubaan	Anita Shabdish	Music & Musicals	Mozez Singh	India	s8807	Movie	March 2 2019
202062	Zubaan	Chittaranjan Tripathy	Dramas	Mozez Singh	India	s8807	Movie	March 2 2019
202063	Zubaan	Chittaranjan Tripathy	International Movies	Mozez Singh	India	s8807	Movie	March 2 2019
202064	Zubaan	Chittaranjan Tripathy	Music & Musicals	Mozez Singh	India	s8807	Movie	March 2 2019

202065 rows × 11 columns

Data analysis

Cast & directors are ignored here because of lot of missing values

q1 - Country vs no_of_movies

In [254]:

```
q1 = df.drop_duplicates(keep='first',subset=['Title'])
q1
```

Out[254]:

	Title	cast	genre	directors	country	show_id	type	date_add
0	Dick Johnson Is Dead	unknown	Documentaries	Kirsten Johnson	United States	s1	Movie	Septemb 25, 20
1	Blood & Water	Ama Qamata	International TV Shows	unknown	South Africa	s2	TV Show	Septemb 24, 20
58	Ganglands	Sami Bouajila	Crime TV Shows	Julien Leclercq	unknown	s3	TV Show	Septemb 24, 20
85	Jailbirds New Orleans	unknown	Docuseries	unknown	unknown	s4	TV Show	Septemb 24, 20
87	Kota Factory	Mayur More	International TV Shows	unknown	India	s5	TV Show	Septemb 24, 20
...	...	...	...	...	...	...	...	
201976	Zodiac	Mark Ruffalo	Cult Movies	David Fincher	United States	s8803	Movie	Novemb 20, 20
202006	Zombie Dumb	unknown	Kids' TV	unknown	unknown	s8804	TV Show	July 1, 20
202009	Zombieland	Jesse Eisenberg	Comedies	Ruben Fleischer	United States	s8805	Movie	Novemb 1, 20
202023	Zoom	Tim Allen	Children & Family Movies	Peter Hewitt	United States	s8806	Movie	January 20
202041	Zubaan	Vicky Kaushal	Dramas	Mozez Singh	India	s8807	Movie	March 20

8807 rows × 11 columns

In [255]:

```python
q1 = q1.groupby('country').agg(count=("country","count")).reset_index()
q1 = q1.sort_values(by='count',ascending=False)
q1 = q1.loc[q1['count']>q1['count'].mean()].head(20)
q1.head(10)
```

Out[255]:

	country	count
80	United States	3211
30	India	1008
86	unknown	831
79	United Kingdom	628
11	Canada	271
37	Japan	259
21	France	212
68	South Korea	211
70	Spain	181
45	Mexico	134

In [256]:

```python
q1 = q1.replace({'country':['United States',' United States']},'US')
q1 = q1.replace({'country':['France',' France']},'France')
q1 = q1.replace({'country':['United Kingdom',' United Kingdom']},'UK')
q1 = q1.replace({'country':['Canada',' Canada']},'canada')
q1.head(10)
```

Out[256]:

	country	count
80	US	3211
30	India	1008
86	unknown	831
79	UK	628
11	canada	271
37	Japan	259
21	France	212
68	South Korea	211
70	Spain	181
45	Mexico	134

In [257]:

```python
q1_final = q1.groupby('country').sum().sort_values(by='count',ascending=False)
q1_final = q1_final.reset_index()
q1_final.head(10)
```

Out[257]:

	country	count
0	US	3211
1	India	1008
2	unknown	831
3	UK	628
4	canada	271
5	Japan	259
6	France	212
7	South Korea	211
8	Spain	181
9	Mexico	134

In [258]:

```python
plt.figure(figsize=(20,8))
sns.barplot(data=q1_final,x='country',y='count')
plt.title("country vs no_of_movies")
plt.show()
```
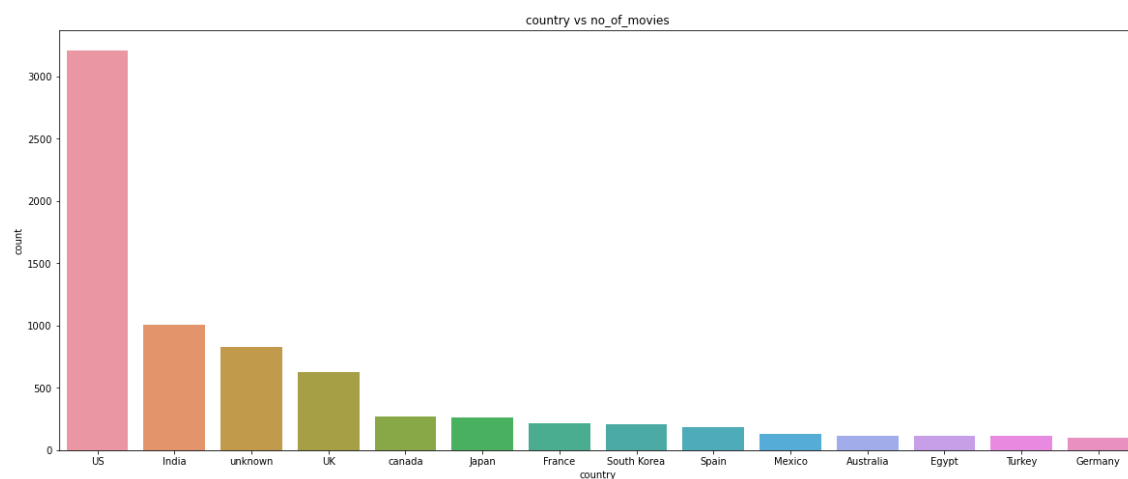
Insight: Above table provides a trend of number of movies/tvshows created so far by each country. Higher number of movies indicates public's fondness towards movies/tvshows

Recommendation: Focus and target top 10 countries from the above graph

q2 - movies vs tv shows

In [276]:

```
q2 = df.drop_duplicates(keep='first',subset=['Title'])
q2
```

Out[276]:

|	Title	cast	genre	directors	country	show_id	type	date_add
0	Dick Johnson Is Dead	unknown	Documentaries	Kirsten Johnson	United States	s1	Movie	Septemb 25, 20
1	Blood & Water	Ama Qamata	International TV Shows	unknown	South Africa	s2	TV Show	Septemb 24, 20
58	Ganglands	Sami Bouajila	Crime TV Shows	Julien Leclercq	unknown	s3	TV Show	Septemb 24, 20
85	Jailbirds New Orleans	unknown	Docuseries	unknown	unknown	s4	TV Show	Septemb 24, 20
87	Kota Factory	Mayur More	International TV Shows	unknown	India	s5	TV Show	Septemb 24, 20
...	...	...	...	...	...	...	...	
201976	Zodiac	Mark Ruffalo	Cult Movies	David Fincher	United States	s8803	Movie	Novemb 20, 20
202006	Zombie Dumb	unknown	Kids' TV	unknown	unknown	s8804	TV Show	July 1, 20
202009	Zombieland	Jesse Eisenberg	Comedies	Ruben Fleischer	United States	s8805	Movie	Novemb 1, 20
202023	Zoom	Tim Allen	Children & Family Movies	Peter Hewitt	United States	s8806	Movie	January 20
202041	Zubaan	Vicky Kaushal	Dramas	Mozez Singh	India	s8807	Movie	March 20

8807 rows × 11 columns

In [286]:

```python
x = df.groupby(['country','type']).agg(count=('country','count')).reset_index()
x = x.sort_values(by=['count','country'],ascending=[False,False])
x = x.head(20)
x
```

Out[286]:

	country	type	count
281	United States	Movie	38316
200	India	Movie	20761
282	United States	TV Show	11552
145	United States	Movie	7501
289	unknown	Movie	6199
290	unknown	TV Show	5698
279	United Kingdom	Movie	5655
213	Japan	TV Show	4673
280	United Kingdom	TV Show	4078
261	South Korea	TV Show	3648
46	France	Movie	3592
187	France	Movie	3015
168	Canada	Movie	2928
143	United Kingdom	Movie	2925
23	Canada	Movie	2810
212	Japan	Movie	2644
263	Spain	Movie	2608
233	Nigeria	Movie	2082
48	Germany	Movie	2032
202	Indonesia	Movie	1988

In [287]:

```python
plt.figure(figsize=(20,8))
sns.barplot(data=x,x='country',y='count',hue='type')
plt.show()
```
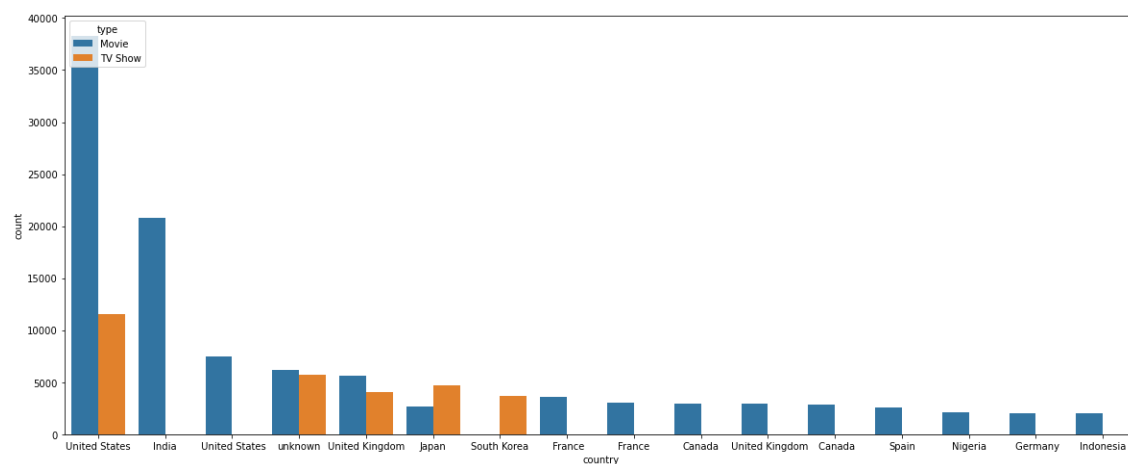
Insight: Provides data regarding countries of what people prefer

Recommendation: Production of movies or tv shows in different countries accordingly

q3 - How has the number of movies released per year changed over the last 20-30 years?

In [337]:

```python
q3 = df.drop_duplicates(keep='first',subset=['Title'])
q3 = q3.loc[q3['type']=='Movie']
q3
```

Out[337]:

|	Title	cast	genre	directors	country	show_id	type	date_a
0	Dick Johnson Is Dead	unknown	Documentaries	Kirsten Johnson	United States	s1	Movie	Septe 25,
159	My Little Pony: A New Generation	Vanessa Hudgens	Children & Family Movies	Robert Cullen	unknown	s7	Movie	Septe 24,
179	Sankofa	Kofi Ghanaba	Dramas	Haile Gerima	United States	s8	Movie	Septe 24,
331	The Starling	Melissa McCarthy	Comedies	Theodore Melfi	United States	s10	Movie	Septe 24,
431	Je Suis Karl	Luna Wedler	Dramas	Christian Schwochow	Germany	s13	Movie	Septe 23,
...	...	...	...	...	...	...	...	
201940	Zinzana	Ali Suliman	Dramas	Majid Al Ansari	United Arab Emirates	s8802	Movie	Mar
201976	Zodiac	Mark Ruffalo	Cult Movies	David Fincher	United States	s8803	Movie	Nove 20,
202009	Zombieland	Jesse Eisenberg	Comedies	Ruben Fleischer	United States	s8805	Movie	Nove 1,
202023	Zoom	Tim Allen	Children & Family Movies	Peter Hewitt	United States	s8806	Movie	Januar
202041	Zubaan	Vicky Kaushal	Dramas	Mozez Singh	India	s8807	Movie	Mar

6131 rows × 11 columns

In [338]:

```python
q3 = pd.DataFrame(q3.groupby('release_year')['Title'].count()).reset_index()
q3 = q3.rename(columns={'Title':'no_of_movies'})
q3 = q3.loc[q3['release_year']>(2021-30)]
q3.head()
```

Out[338]:

	release_year	no_of_movies
43	1992	20
44	1993	24
45	1994	20
46	1995	23
47	1996	21

In [334]:

```python
plt.figure(figsize=(20,8))
sns.barplot(data=q3,x='release_year',y='no_of_movies')
plt.title("year vs no_of_movies")
plt.show()
```
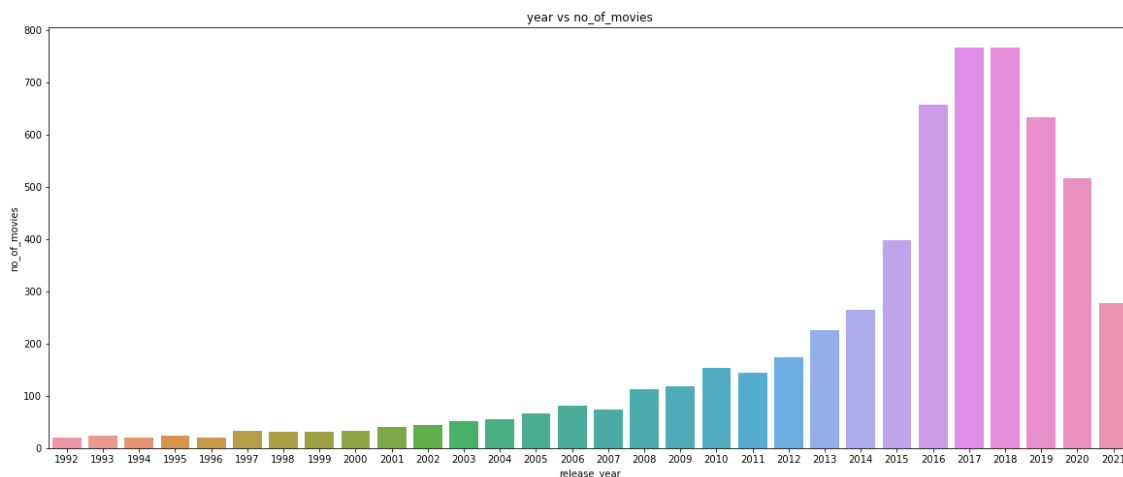
Insight: provides data about the number of movies created year on year. Creation of movies has peaked in 2017-2018 and since then it has reduced by more than 50%

Recommendation: Lower trend in creation of movies in recent times indicates people are preferring TV shows

year vs tv shows

In [352]:

```python
q4 = df.drop_duplicates(keep='first',subset=['Title'])
q4 = q4.loc[q4['type']=='TV Show']
q4.head()
```

Out[352]:

	Title	cast	genre	directors	country	show_id	type	date_added	rele
1	Blood & Water	Ama Qamata	International TV Shows	unknown	South Africa	s2	TV Show	September 24, 2021	
58	Ganglands	Sami Bouajila	Crime TV Shows	Julien Leclercq	unknown	s3	TV Show	September 24, 2021	
85	Jailbirds New Orleans	unknown	Docuseries	unknown	unknown	s4	TV Show	September 24, 2021	
87	Kota Factory	Mayur More	International TV Shows	unknown	India	s5	TV Show	September 24, 2021	
111	Midnight Mass	Kate Siegel	TV Dramas	Mike Flanagan	unknown	s6	TV Show	September 24, 2021	

In [353]:

```python
q4 = pd.DataFrame(q4.groupby('release_year')['Title'].count()).reset_index()
q4 = q4.rename(columns={'Title':'no_of_tv_shows'})
q4 = q4.loc[q4['release_year']>(2021-30)]
q4.head()
```
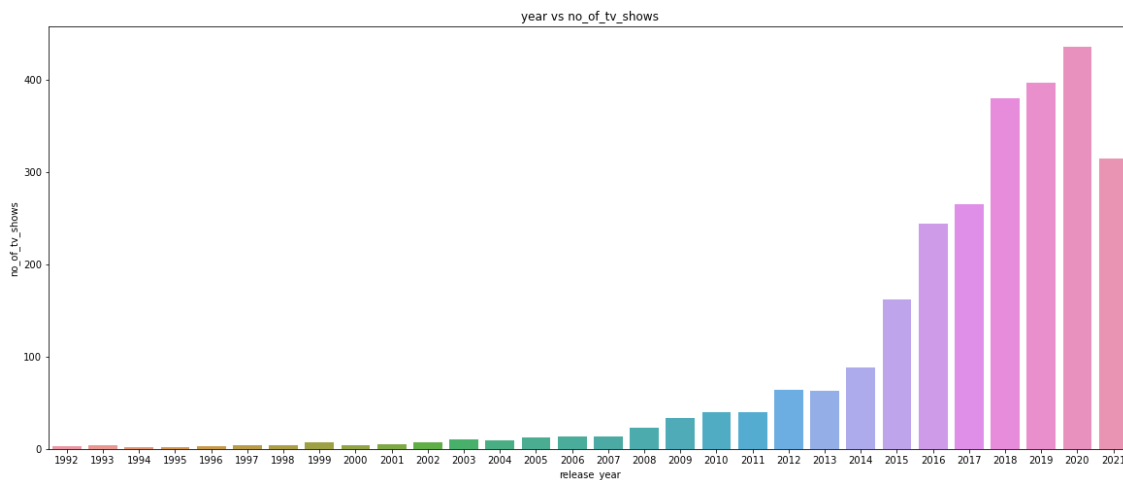
Out[353]:

	release_year	no_of_tv_shows
16	1992	3
17	1993	4
18	1994	2
19	1995	2
20	1996	3

In [355]:

```python
plt.figure(figsize=(20,8))
sns.barplot(data=q4,x='release_year',y='no_of_tv_shows')
plt.title("year vs no_of_tv_shows")
plt.show()
```


Insight: provides data about the number of TV shows created year on year. Creation of TV shows has peaked in 2020.

Recommendation: Lower trend in movies & upper trend in TV shows indicates that people prefer TV shows over movies

q5 - year vs genre in united states

In [358]:

```python
df['genre'].nunique()
```

Out[358]:

73

In [410]:

```
q5 = df.drop_duplicates(keep='first',subset='Title')
q5.head()
```

Out[410]:

	Title	cast	genre	directors	country	show_id	type	date_added	re
0	Dick Johnson Is Dead	unknown	Documentaries	Kirsten Johnson	United States	s1	Movie	September 25, 2021	
1	Blood & Water	Ama Qamata	International TV Shows	unknown	South Africa	s2	TV Show	September 24, 2021	
58	Ganglands	Sami Bouajila	Crime TV Shows	Julien Leclercq	unknown	s3	TV Show	September 24, 2021	
85	Jailbirds New Orleans	unknown	Docuseries	unknown	unknown	s4	TV Show	September 24, 2021	
87	Kota Factory	Mayur More	International TV Shows	unknown	India	s5	TV Show	September 24, 2021	

In [411]:

```
q5 = q5.groupby(['release_year','genre']).agg(count=('Title','count'))
q5 = q5.reset_index()
q5 = q5.sort_values(by=['release_year','count'],ascending=[True,False])
q5 = q5.drop_duplicates(keep='first',subset='release_year')
q5.head()
```

Out[411]:

	release_year	genre	count
0	1925	TV Shows	1
1	1942	Classic Movies	2
3	1943	Documentaries	2
4	1944	Classic Movies	3
5	1945	Classic Movies	3

In [409]:

```python
plt.figure(figsize=(20,8))
sns.scatterplot(data=q5,x='release_year',y='genre')
plt.title("year vs genre")
plt.show()
```
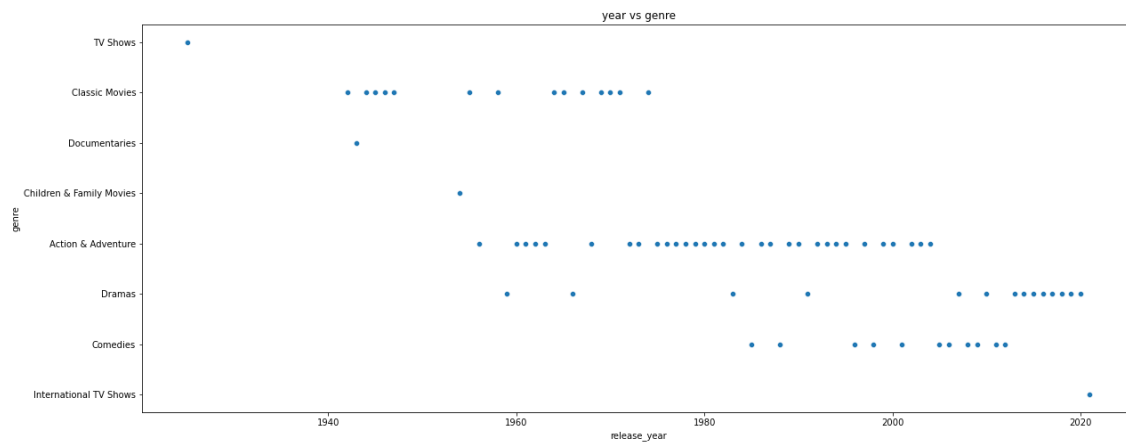

Insight: This graph provides data regarding genres of movies over a period of time. People are shifting from classical movies to drama/comedies/action

Recommendation: Production / listing of movies/TV shows on Netflix should be from genre drama/comedies/action

q6 - year vs duration

In [509]:

```python
q6 = df.drop_duplicates(keep='first',subset='Title')
q6 = q6.loc[q6['type']=="Movie"]
q6.head()
```

Out[509]:

	Title	cast	genre	directors	country	show_id	type	date_added
0	Dick Johnson Is Dead	unknown	Documentaries	Kirsten Johnson	United States	s1	Movie	September 25, 2021
159	My Little Pony: A New Generation	Vanessa Hudgens	Children & Family Movies	Robert Cullen	unknown	s7	Movie	September 24, 2021
179	Sankofa	Kofi Ghanaba	Dramas	Haile Gerima	United States	s8	Movie	September 24, 2021
331	The Starling	Melissa McCarthy	Comedies	Theodore Melfi	United States	s10	Movie	September 24, 2021
431	Je Suis Karl	Luna Wedler	Dramas	Christian Schwochow	Germany	s13	Movie	September 23, 2021

In [510]:

```python
x = q6['duration'].str.split(expand=True)
x = x.fillna(0)
x.isna().sum()
x = x.astype({0:'int64'})
x.head()
```

Out[510]:

	0	1
0	90	min
159	91	min
179	125	min
331	104	min
431	127	min

In [511]:

```python
q6['duration']=x[0]
q6.head()
```

Out[511]:

|	Title	cast	genre	directors	country	show_id	type	date_added
0	Dick Johnson Is Dead	unknown	Documentaries	Kirsten Johnson	United States	s1	Movie	September 25, 2021
159	My Little Pony: A New Generation	Vanessa Hudgens	Children & Family Movies	Robert Cullen	unknown	s7	Movie	September 24, 2021
179	Sankofa	Kofi Ghanaba	Dramas	Haile Gerima	United States	s8	Movie	September 24, 2021
331	The Starling	Melissa McCarthy	Comedies	Theodore Melfi	United States	s10	Movie	September 24, 2021
431	Je Suis Karl	Luna Wedler	Dramas	Christian Schwochow	Germany	s13	Movie	September 23, 2021

In [512]:

```python
q6 = q6.groupby('release_year').agg(mean_dur=('duration','mean')).reset_index()
q6.head()
```

Out[512]:

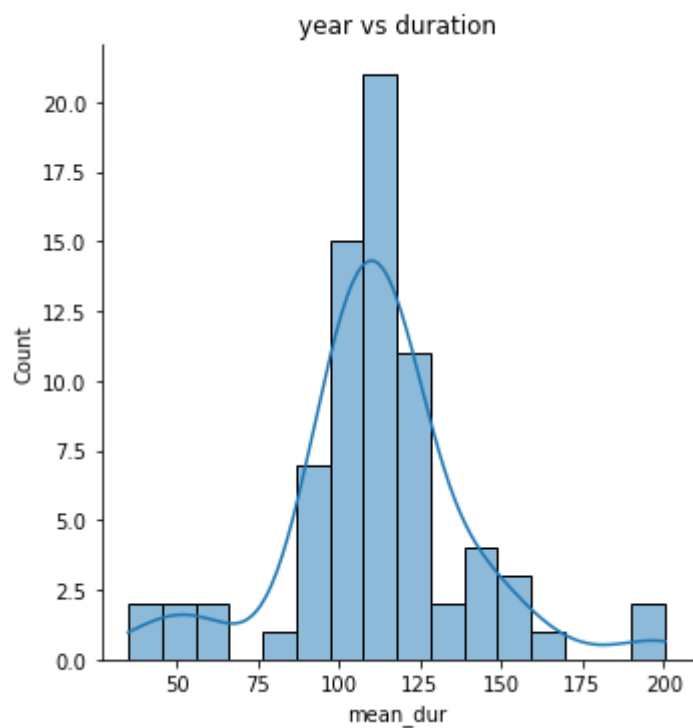|	release_year	mean_dur
0	1942	35.000000
1	1943	62.666667
2	1944	52.000000
3	1945	51.333333
4	1946	58.000000

In [498]:

```python
plt.figure(figsize=(20,8))
sns.barplot(data=q6,x='release_year',y='mean_dur')
plt.title("year vs duration")
plt.show()
```
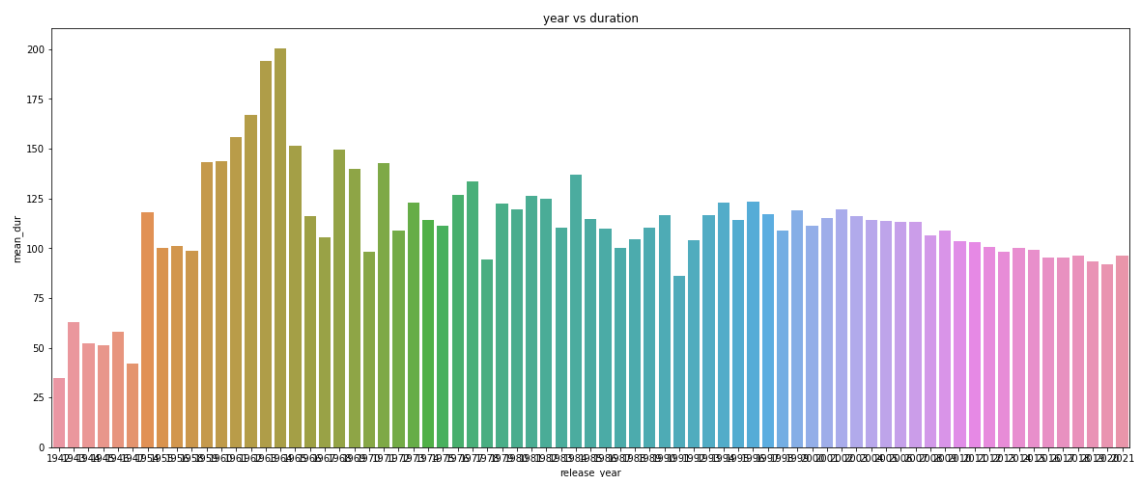
In [489]:

```python
plt.figure(figsize=(20,8))
sns.displot(data=q6,x='mean_dur',kde=True)
plt.title("year vs duration")
plt.show()
```

<Figure size 1440x576 with 0 Axes>

Insight: Above trend shows the duration of movies over a period of time.

Recommendation: Since people prefer duration of around 110 mins for a movie, movies of that duration should be created. Reduction in duration time also indicates people are preferring shorter movies or TV shows for that matter
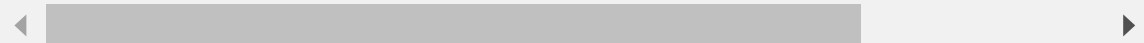
q7 - rating vs movies

In [517]:

```
q7 = df.drop_duplicates(keep='first',subset='Title')
q7.head()
```

Out[517]:

	Title	cast	genre	directors	country	show_id	type	date_added	re
0	Dick Johnson Is Dead	unknown	Documentaries	Kirsten Johnson	United States	s1	Movie	September 25, 2021	
1	Blood & Water	Ama Qamata	International TV Shows	unknown	South Africa	s2	TV Show	September 24, 2021	
58	Ganglands	Sami Bouajila	Crime TV Shows	Julien Leclercq	unknown	s3	TV Show	September 24, 2021	
85	Jailbirds New Orleans	unknown	Docuseries	unknown	unknown	s4	TV Show	September 24, 2021	
87	Kota Factory	Mayur More	International TV Shows	unknown	India	s5	TV Show	September 24, 2021	

In [520]:

```python
q7 = q7.groupby('rating').agg(count=('Title','count')).reset_index()
q7
```

Out[520]:

	rating	count
0	66 min	1
1	74 min	1
2	84 min	1
3	G	41
4	NC-17	3
5	NR	80
6	PG	287
7	PG-13	490
8	R	799
9	TV-14	2160
10	TV-G	220
11	TV-MA	3207
12	TV-PG	863
13	TV-Y	307
14	TV-Y7	334
15	TV-Y7-FV	6
16	UR	3

In [522]:

```python
plt.figure(figsize=(12,8))
sns.barplot(data=q7,x='rating',y='count')
plt.title("year vs duration")
plt.show()
```
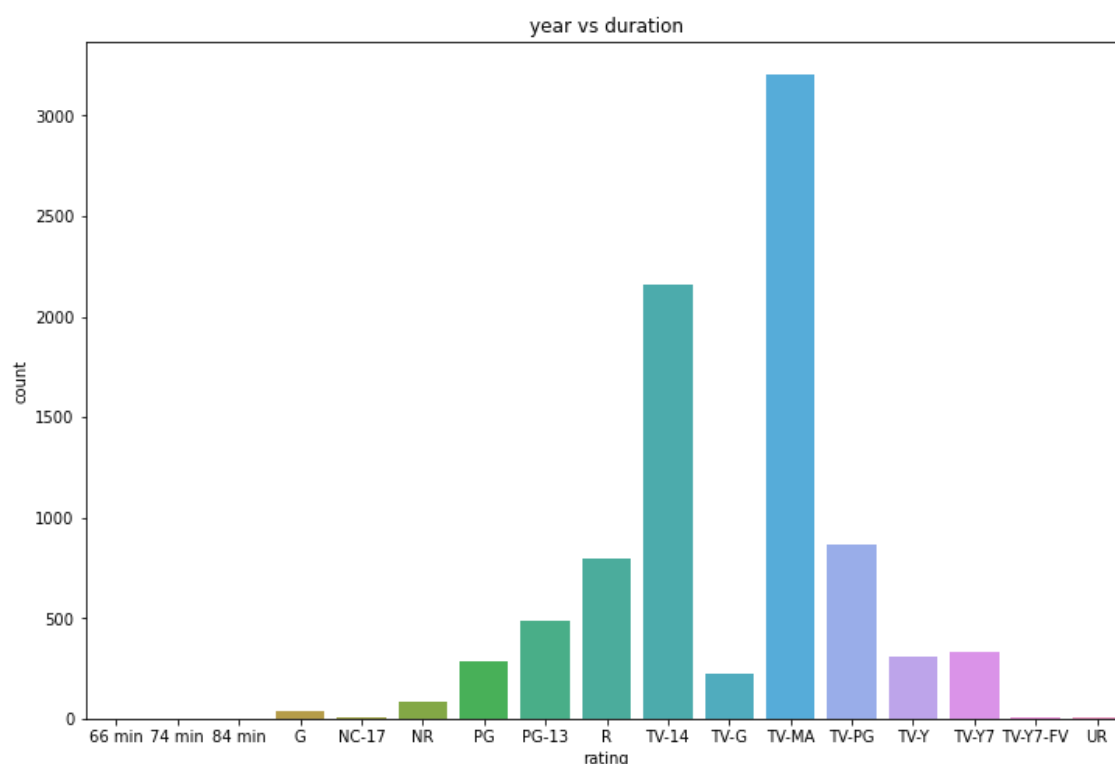

Insight: above data provides information regarding rating

Recommendation: Movies ratings with higher number of movies is preferrable TV-14/TV-MA for a success

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: