OTT Streaming Platform Business case analysis

The dataset is extracted from one of the most popular online video and media streaming services. They have over 10000 movies or tv shows available on their platform, as of mid-2021, they have over 222M Subscribers globally.

This tabular dataset consists of listings of all the movies and tv shows available on the popular OTT platform **Netflix**, along with details such as - cast, directors, ratings, release year, duration, etc.

Downloading the necessary packages

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Importing the dataset

In [2]:
```
!gdown 1cnenKcHofZDvIUVpVMekDIn9nrd8pWv7
```

```
Downloading...
From: https://drive.google.com/uc?id=1cnenKcHofZDvIUVpVMekDIn9nrd8pWv7 (https://drive.google.com/uc?id=1cnenKcHofZDvIUVpVMekDIn9nrd8pWv7)
To: /content/netflix.csv
100% 3.40M/3.40M [00:00<00:00, 94.9MB/s]
```

In [3]:
```
df = pd.read_csv('/content/netflix.csv')
df.head()
```

Out[3]:

	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...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train I...

Initial analaysis: We have to analyze the dataset extracted from the OTT streaming platform to come up with business insights which would help the company to develop their business globally.

Shape of the dataset

In [4]:
```
df.shape
```

Out[4]:
```
(8807, 12)
```

Length of the raw dataset

In [5]:
```
len(df)
```

Out[5]:
```
8807
```

There are about 8807 rows and about 12 columns in the dataset.

Data type that is defined for each of the feature in the dataframe.

In [6]:
```
df.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
```

From the above we can see that all the features are of data type string except the release year feature which is of type integer.

In [7]:
```
df.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

Since we have only one numerical feature in the dataset, we will be getting the statistical metrics of release year feature.

If we want to include all the features, then we have to pass the includeAll parameter to the method.

In [8]:
```
df.describe(include="all")
```

Out[8]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
count	8807	8807	8807	6173	7982	7976	8797	8807.000000	8803	8804	8807	8807
unique	8807	2	8807	4528	7692	748	1767	NaN	17	220	514	8775
top	s1	Movie	Dick Johnson Is Dead	Rajiv Chilaka	David Attenborough	United States	January 1, 2020	NaN	TV-MA	1 Season	Dramas, International Movies	Paranormal activity at a lush, abandoned prope...
freq	1	6131	1	19	19	2818	109	NaN	3207	1793	362	4
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2014.180198	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	8.819312	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1925.000000	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2013.000000	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2017.000000	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2019.000000	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2021.000000	NaN	NaN	NaN	NaN

We can see some extra information got extracted from the dataset like total number of Unique values in each of the defined columns, Top most record of each of the column in the dataframe, total number of records (excluding the NAN values) for each of the column in the dataframe.

Missing values present in the dataset:

In [9]:
```
df.isna()
```

Out[9]:

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

8807 rows × 12 columns

We can see there are many missing values here and there in the dataset.

In [10]:
```
df.isna().sum()
```

Out[10]:
```
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
```

From the above we will be getting the total number of missing values in each feature column of the dataset.

We can see there are 2634 missing values in director column, 825 in cast column, 831 in country, 10 in date_added, 4 in rating and 3 in duration.

Columns in the dataset:

In [11]:
```
df.columns
```

Out[11]:
```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object')
```

Total number of unique values in the dataset.

In [12]:
```
df.nunique()
```

Out[12]:
```
show_id       8807
type             2
title         8807
director      4528
cast          7692
country        748
date_added    1767
release_year    74
rating          17
duration       220
listed_in      514
description   8775
dtype: int64
```

For each of the columns in the dataset we can see the total number of unique values in them.

For each of the feature in the dataframe, lets explore the number of occurences of each value in that column.

In [13]:
```
for col in df.columns:
  print("Number of occurences of values in {} column".format(col))
  print(df[col].value_counts())
  print('\n')
```

```
Number of occurences of values in show_id column
s1       1
s5875    1
s5869    1
s5870    1
s5871    1
        ..
s2931    1
s2930    1
s2929    1
s2928    1
s8807    1
Name: show_id, Length: 8807, dtype: int64


Number of occurences of values in type column
Movie      6131
TV Show    2676
Name: type, dtype: int64
```

Pre-processing of the data

In [14]:
```
df.head()
```

Out[14]:

	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...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train I...

Lets explore one record from the dataframe to understand how untidy the data is:

In [15]:
```
df.iloc[2,:]['cast']
```

Out[15]:
```
'Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabiha Akkari, Sofia Lesaffre, Salim Kechiouche, Noureddine Farihi, Geert
Van Rampelberg, Bakary Diombera'
```

In [16]:
```
df.iloc[2,:]['cast']
```

Out[16]:
```
'Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabiha Akkari, Sofia Lesaffre, Salim Kechiouche, Noureddine Farihi, Geert
Van Rampelberg, Bakary Diombera'
```

We can see there are multiple cast members mentioned in the single record of cast and Multiple categories mentioned in the listed_in feature of the dataset.

We have to un-nest them to proceed further with the dataset.

In [17]:
```
df.head()
```

Out[17]:

	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...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train I...

Starting with the cast feature

From the above we can see that for some of the movies there are about 50 directors and thats why there are about 50 columns got created.

In [18]:
```
df.loc[df['cast'].str.split(',',expand=True)[49].isin([' Lovie Simone', ' Rafe Spall'])]['cast']
```

Out[18]:
```
1854    Danielle Brooks, Oscar Nuñez, Mike Colter, Hea...
3774    Jesse Plemons, Cristin Milioti, Jimmi Simpson,...
Name: cast, dtype: object
```

The above 2 are the movies that have about 50 directors. Thus we have to separate out these cast members into separate records and continue the analysis further.

In [19]:
```
data = df.copy(deep=True)   #Creating restore point for the dataset.
```

We can see there are comma separated values in the cast records. Thus we will split them based on ',' (comma) separator and then convert them into list and store them in the same place.

Using lambda function and .apply() method, we can convert the string data into list of comma separated values.

In [20]:
```
df['cast'] = df['cast'].apply(lambda x : list(x.split(',')) if type(x)==str else x)
df
```

Out[20]:

	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, Th...	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...
2	s3	TV Show	Ganglands	Julien Leclercq	[Sami Bouajila, Tracy Gotoas, Samuel Jouy, ...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	[Mayur More, Jitendra Kumar, Ranjan Raj, Al...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train I...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	[Mark Ruffalo, Jake Gyllenhaal, Robert Downe...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie	Zombieland	Ruben Fleischer	[Jesse Eisenberg, Woody Harrelson, Emma Ston...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...
8805	s8806	Movie	Zoom	Peter Hewitt	[Tim Allen, Courteney Cox, Chevy Chase, Kat...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie	Zubaan	Mozez Singh	[Vicky Kaushal, Sarah-Jane Dias, Raaghav Cha...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...

8807 rows × 12 columns

In [21]:
```
data = df.copy(deep=True)   #Creating restore point for the dataset.
```

Example of cast details for one of the TV show.

In [22]:
```
df.iloc[2,:]['cast']
```

Out[22]:
```
['Sami Bouajila',
 ' Tracy Gotoas',
 ' Samuel Jouy',
 ' Nabiha Akkari',
 ' Sofia Lesaffre',
 ' Salim Kechiouche',
 ' Noureddine Farihi',
 ' Geert Van Rampelberg',
 ' Bakary Diombera']
```

With the help of .explode() method we can explode the list of cast details into individual records which thus creates a unique director, actor pair.

In [23]:
```
exp_df = df.explode('cast')
exp_df
```

Out[23]:

	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	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...
1	s2	TV Show	Blood & Water	NaN	Khosi Ngema	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...
1	s2	TV Show	Blood & Water	NaN	Gail Mabalane	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...
1	s2	TV Show	Blood & Water	NaN	Thabang Molaba	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...
...	...	...	...	...	...	...	...	...	...	...	...	...
8806	s8807	Movie	Zubaan	Mozez Singh	Manish Chaudhary	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...
8806	s8807	Movie	Zubaan	Mozez Singh	Meghna Malik	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...
8806	s8807	Movie	Zubaan	Mozez Singh	Malkeet Rauni	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...

64951 rows × 12 columns

We can see the number of records have increased in the exp_df when compared to the original dataset.

In [24]:
```
df.loc[df['show_id']=='s2']['cast'].values
```

Out[24]:
```
array([list(['Ama Qamata', ' Khosi Ngema', ' Gail Mabalane', ' Thabang Molaba', ' Dillon Windvogel', ' Natasha Thahan
e', ' Arno Greeff', ' Xolile Tshabalala', ' Getmore Sithole', ' Cindy Mahlangu', ' Ryle De Morny', ' Greteli Fincham
', ' Sello Maake Ka-Ncube', ' Odwa Gwanya', ' Mekaila Mathys', ' Sandi Schultz', ' Duane Williams', ' Shamilla Miller
', ' Patrick Mofokeng'])],
      dtype=object)
```

In [25]:
```
exp_df.loc[exp_df['show_id']=='s2']
```

Out[25]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	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...
1	s2	TV Show	Blood & Water	NaN	Khosi Ngema	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...
1	s2	TV Show	Blood & Water	NaN	Gail Mabalane	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...
1	s2	TV Show	Blood & Water	NaN	Thabang Molaba	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...
1	s2	TV Show	Blood & Water	NaN	Dillon Windvogel	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...
1	s2	TV Show	Blood & Water	NaN	Natasha Thahane	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...
1	s2	TV Show	Blood & Water	NaN	Arno Greeff	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...
1	s2	TV Show	Blood & Water	NaN	Xolile Tshabalala	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...
1	s2	TV Show	Blood & Water	NaN	Getmore Sithole	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...
1	s2	TV Show	Blood & Water	NaN	Cindy Mahlangu	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...
1	s2	TV Show	Blood & Water	NaN	Ryle De Morny	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...
1	s2	TV Show	Blood & Water	NaN	Greteli Fincham	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...
1	s2	TV Show	Blood & Water	NaN	Sello Maake Ka-Ncube	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...
1	s2	TV Show	Blood & Water	NaN	Odwa Gwanya	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...
1	s2	TV Show	Blood & Water	NaN	Mekaila Mathys	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...
1	s2	TV Show	Blood & Water	NaN	Sandi Schultz	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...
1	s2	TV Show	Blood & Water	NaN	Duane Williams	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...
1	s2	TV Show	Blood & Water	NaN	Shamilla Miller	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...
1	s2	TV Show	Blood & Water	NaN	Patrick Mofokeng	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...

Similar to Cast feature of the dataframe, there are other features like Director, Country and listed_in which has to be un-nested in the dataset.

We will apply the same split and explode method for these features also.

In [26]:
```
exp_df['director'] = exp_df['director'].apply(lambda x:list(x.split(', ')) if type(x)==str else x)   #Splitting the st
exp_df['country'] = exp_df['country'].apply(lambda x:list(x.split(', ')) if type(x)==str else x)
exp_df['listed_in'] = exp_df['listed_in'].apply(lambda x:list(x.split(', ')) if type(x)==str else x)
```

Exploding the features one by one:

In [27]:
```
exp_df = exp_df.explode("director")
exp_df = exp_df.explode("country")
exp_df = exp_df.explode("listed_in")
exp_df.reset_index(drop=True,inplace=True)
exp_df
```

Out[27]:

	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	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...
3	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Mysteries	After crossing paths at a party, a Cape Town t...
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
...	...	...	...	...	...	...	...	...	...	...	...	...
201986	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201987	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...
201988	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Dramas	A scrappy but poor boy worms his way into a ty...
201989	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201990	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...

201991 rows × 12 columns

Lets check the Unique values in the country feature.

In [28]:
```
exp_df['country'].unique()
```

Out[28]:
```
array(['United States', 'South Africa', nan, 'India', 'Ghana',
       'Burkina Faso', 'United Kingdom', 'Germany', 'Ethiopia',
       'Czech Republic', 'Mexico', 'Turkey', 'Australia', 'France',
       'Finland', 'China', 'Canada', 'Japan', 'Nigeria', 'Spain',
       'Belgium', 'South Korea', 'Singapore', 'Italy', 'Romania',
       'Argentina', 'Venezuela', 'Hong Kong', 'Russia', '', 'Ireland',
       'Nepal', 'New Zealand', 'Brazil', 'Greece', 'Jordan', 'Colombia',
       'Switzerland', 'Israel', 'Taiwan', 'Bulgaria', 'Algeria', 'Poland',
       'Saudi Arabia', 'Thailand', 'Indonesia', 'Egypt', 'Denmark',
       'Kuwait', 'Netherlands', 'Malaysia', 'Vietnam', 'Hungary',
       'Sweden', 'Lebanon', 'Syria', 'Philippines', 'Iceland',
       'United Arab Emirates', 'Norway', 'Qatar', 'Mauritius', 'Austria',
       'Cameroon', 'Palestine', 'Uruguay', 'United Kingdom,', 'Kenya',
       'Chile', 'Luxembourg', 'Cambodia', 'Bangladesh', 'Portugal',
       'Cayman Islands', 'Senegal', 'Serbia', 'Malta', 'Namibia',
       'Angola', 'Peru', 'Mozambique', 'Cambodia,', 'Belarus', 'Zimbabwe',
       'Puerto Rico', 'Pakistan', 'Cyprus', 'Guatemala', 'Iraq', 'Malawi',
       'Paraguay', 'Croatia', 'Iran', 'West Germany', 'United States,',
       'Albania', 'Georgia', 'Soviet Union', 'Morocco', 'Slovakia',
       'Ukraine', 'Bermuda', 'Ecuador', 'Armenia', 'Mongolia', 'Bahamas',
       'Sri Lanka', 'Latvia', 'Liechtenstein', 'Cuba', 'Nicaragua',
       'Poland,', 'Slovenia', 'Dominican Republic', 'Samoa', 'Azerbaijan',
       'Botswana', 'Vatican City', 'Jamaica', 'Kazakhstan', 'Lithuania',
       'Afghanistan', 'Somalia', 'Sudan', 'Panama', 'Uganda',
       'East Germany', 'Montenegro'], dtype=object)
```

In [29]:
```
exp_df['country'].nunique()
```

Out[29]:
```
127
```

From the above we can see there is some white-spaces that are added at the start of country names.

So we have to strip out the empty spaces in the start and end of country feature.

In [30]:
```
exp_df['country'] = exp_df['country'].str.strip(' ,')
exp_df['country'].nunique()
```

Out[30]:
```
123
```

In [31]:
```
exp_df.loc[exp_df['country']=='']
```

Out[31]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
4737	s194	TV Show	D.P.	NaN	Jung Hae-in		August 27, 2021	2021	TV-MA	1 Season	International TV Shows	A young private's assignment to capture army d...
4738	s194	TV Show	D.P.	NaN	Jung Hae-in		August 27, 2021	2021	TV-MA	1 Season	TV Dramas	A young private's assignment to capture army d...
4741	s194	TV Show	D.P.	NaN	Koo Kyo-hwan		August 27, 2021	2021	TV-MA	1 Season	International TV Shows	A young private's assignment to capture army d...
4742	s194	TV Show	D.P.	NaN	Koo Kyo-hwan		August 27, 2021	2021	TV-MA	1 Season	TV Dramas	A young private's assignment to capture army d...
4745	s194	TV Show	D.P.	NaN	Kim Sung-kyun		August 27, 2021	2021	TV-MA	1 Season	International TV Shows	A young private's assignment to capture army d...
4746	s194	TV Show	D.P.	NaN	Kim Sung-kyun		August 27, 2021	2021	TV-MA	1 Season	TV Dramas	A young private's assignment to capture army d...
4749	s194	TV Show	D.P.	NaN	Son Suk-ku		August 27, 2021	2021	TV-MA	1 Season	International TV Shows	A young private's assignment to capture army d...
4750	s194	TV Show	D.P.	NaN	Son Suk-ku		August 27, 2021	2021	TV-MA	1 Season	TV Dramas	A young private's assignment to capture army d...
9056	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga		July 30, 2021	2014	TV-14	103 min	Dramas	After a decade in prison, a Palestinian man wi...
9057	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga		July 30, 2021	2014	TV-14	103 min	Independent Movies	After a decade in prison, a Palestinian man wi...
9058	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga		July 30, 2021	2014	TV-14	103 min	International Movies	After a decade in prison, a Palestinian man wi...
9065	s366	Movie	Eyes of a Thief	Najwa Najjar	Souad Massi		July 30, 2021	2014	TV-14	103 min	Dramas	After a decade in prison, a Palestinian man wi...
9066	s366	Movie	Eyes of a Thief	Najwa Najjar	Souad Massi		July 30, 2021	2014	TV-14	103 min	Independent Movies	After a decade in prison, a Palestinian man wi...
9067	s366	Movie	Eyes of a Thief	Najwa Najjar	Souad Massi		July 30, 2021	2014	TV-14	103 min	International Movies	After a decade in prison, a Palestinian man wi...
9074	s366	Movie	Eyes of a Thief	Najwa Najjar	Suhail Haddad		July 30, 2021	2014	TV-14	103 min	Dramas	After a decade in prison, a Palestinian man wi...
9075	s366	Movie	Eyes of a Thief	Najwa Najjar	Suhail Haddad		July 30, 2021	2014	TV-14	103 min	Independent Movies	After a decade in prison, a Palestinian man wi...
9076	s366	Movie	Eyes of a Thief	Najwa Najjar	Suhail Haddad		July 30, 2021	2014	TV-14	103 min	International Movies	After a decade in prison, a Palestinian man wi...
9083	s366	Movie	Eyes of a Thief	Najwa Najjar	Malak Ermileh		July 30, 2021	2014	TV-14	103 min	Dramas	After a decade in prison, a Palestinian man wi...
9084	s366	Movie	Eyes of a Thief	Najwa Najjar	Malak Ermileh		July 30, 2021	2014	TV-14	103 min	Independent Movies	After a decade in prison, a Palestinian man wi...
9085	s366	Movie	Eyes of a Thief	Najwa Najjar	Malak Ermileh		July 30, 2021	2014	TV-14	103 min	International Movies	After a decade in prison, a Palestinian man wi...
9092	s366	Movie	Eyes of a Thief	Najwa Najjar	Maisa Abd Elhadi		July 30, 2021	2014	TV-14	103 min	Dramas	After a decade in prison, a Palestinian man wi...
9093	s366	Movie	Eyes of a Thief	Najwa Najjar	Maisa Abd Elhadi		July 30, 2021	2014	TV-14	103 min	Independent Movies	After a decade in prison, a Palestinian man wi...
9094	s366	Movie	Eyes of a Thief	Najwa Najjar	Maisa Abd Elhadi		July 30, 2021	2014	TV-14	103 min	International Movies	After a decade in prison, a Palestinian man wi...
9101	s366	Movie	Eyes of a Thief	Najwa Najjar	Walid Abdul Salam		July 30, 2021	2014	TV-14	103 min	Dramas	After a decade in prison, a Palestinian man wi...
9102	s366	Movie	Eyes of a Thief	Najwa Najjar	Walid Abdul Salam		July 30, 2021	2014	TV-14	103 min	Independent Movies	After a decade in prison, a Palestinian man wi...
9103	s366	Movie	Eyes of a Thief	Najwa Najjar	Walid Abdul Salam		July 30, 2021	2014	TV-14	103 min	International Movies	After a decade in prison, a Palestinian man wi...

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
9110	s366	Movie	Eyes of a Thief	Najwa Najjar	Nisreen Faour		July 30, 2021	2014	TV-14	103 min	Dramas	After a decade in prison, a Palestinian man wi...
9111	s366	Movie	Eyes of a Thief	Najwa Najjar	Nisreen Faour		July 30, 2021	2014	TV-14	103 min	Independent Movies	After a decade in prison, a Palestinian man wi...
9112	s366	Movie	Eyes of a Thief	Najwa Najjar	Nisreen Faour		July 30, 2021	2014	TV-14	103 min	International Movies	After a decade in prison, a Palestinian man wi...
9119	s366	Movie	Eyes of a Thief	Najwa Najjar	Areen Omari		July 30, 2021	2014	TV-14	103 min	Dramas	After a decade in prison, a Palestinian man wi...
9120	s366	Movie	Eyes of a Thief	Najwa Najjar	Areen Omari		July 30, 2021	2014	TV-14	103 min	Independent Movies	After a decade in prison, a Palestinian man wi...

In [32]:
```
exp_df.loc[exp_df['cast']=='Khaled Abol El Naga']
```

Out[32]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
9056	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga		July 30, 2021	2014	TV-14	103 min	Dramas	After a decade in prison, a Palestinian man wi...
9057	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga		July 30, 2021	2014	TV-14	103 min	Independent Movies	After a decade in prison, a Palestinian man wi...
9058	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga		July 30, 2021	2014	TV-14	103 min	International Movies	After a decade in prison, a Palestinian man wi...
9059	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga	France	July 30, 2021	2014	TV-14	103 min	Dramas	After a decade in prison, a Palestinian man wi...
9060	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga	France	July 30, 2021	2014	TV-14	103 min	Independent Movies	After a decade in prison, a Palestinian man wi...
9061	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga	France	July 30, 2021	2014	TV-14	103 min	International Movies	After a decade in prison, a Palestinian man wi...
9062	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga	Algeria	July 30, 2021	2014	TV-14	103 min	Dramas	After a decade in prison, a Palestinian man wi...
9063	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga	Algeria	July 30, 2021	2014	TV-14	103 min	Independent Movies	After a decade in prison, a Palestinian man wi...
9064	s366	Movie	Eyes of a Thief	Najwa Najjar	Khaled Abol El Naga	Algeria	July 30, 2021	2014	TV-14	103 min	International Movies	After a decade in prison, a Palestinian man wi...

As we can see from the above we can see there are few records with empty countries even though they have some country values.

These got created as a result of explode of feature.

Lets remove these records from the dataset.

In [33]:
```
exp_df.loc[exp_df['country']==''].index
```

Out[33]:
```
Int64Index([4737, 4738, 4741, 4742, 4745, 4746, 4749, 4750, 9056, 9057, 9058,
            9065, 9066, 9067, 9074, 9075, 9076, 9083, 9084, 9085, 9092, 9093,
            9094, 9101, 9102, 9103, 9110, 9111, 9112, 9119, 9120, 9121],
           dtype='int64')
```

In [34]:
```
exp_df.drop(exp_df.loc[exp_df['country']==''].index,axis=0)
```

Out[34]:

	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	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...
3	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Mysteries	After crossing paths at a party, a Cape Town t...
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
...	...	...	...	...	...	...	...	...	...	...	...	...
201986	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201987	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...
201988	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Dramas	A scrappy but poor boy worms his way into a ty...
201989	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201990	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...

201959 rows × 12 columns

Similarly lets strip away the white spaces from other exploded columns like cast, director and listed_in.

In [35]:
```
exp_df['cast'].nunique()  #Number of unique cast before stripping away the white spaces and comma's.
```

Out[35]: 39296

In [36]:
```
exp_df['cast'] = exp_df['cast'].str.strip(' ,')
exp_df['director'] = exp_df['director'].str.strip(' ,')
exp_df['listed_in'] = exp_df['listed_in'].str.strip(' ,')
```

In [37]:
```
exp_df['cast'].nunique()  #Number of unique cast after stripping away the white spaces and comma's.
```

Out[37]: 36439

The shape of exploded data

In [38]:
```
exp_df.shape
```

Out[38]: (201991, 12)

We can see all the features are now un-nested and the number of rows has been increased significantly in the exploded dataframe compared to the original dataframe.

In [39]:
```
#The number of missing values in the exploded dataset
exp_df.isna().sum()
```

Out[39]:
```
show_id            0
type               0
title              0
director       50643
cast            2146
country        11897
date_added       158
release_year       0
rating            67
duration           3
listed_in          0
description        0
dtype: int64
```

As the description feature of the dataset cannot be preprocessed without the use of NLP, dropping the feature for simplicity will cause no harm.

In [40]:
```
exp_df.drop('description',axis=1,inplace=True)
exp_df
```

Out[40]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Dramas
3	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Mysteries
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows
...	...	...	...	...	...	...	...	...	...	...	...
201986	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	International Movies
201987	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals
201988	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Dramas
201989	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	International Movies
201990	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals

201991 rows × 11 columns

Savepoint for exploded dataset:

In [41]:
```
exp_data = exp_df.copy(deep=True)
```

The Popular Director and Actor combo that is popular in the platform.

In [42]:
```
exp_df.groupby(['director','cast'])['title'].nunique().sort_values(ascending=False)
```

Out[42]:
```
director        cast
Rajiv Chilaka   Julie Tejwani          19
                Rajesh Kava            19
                Jigna Bhardwaj         18
                Rupa Bhimani           18
                Vatsal Dubey           16
                                       ..
Huang Jianming  Anthony Padilla         1
                Chevy Chase             1
                Ian Hecox               1
                Jenn McAllister         1
Şenol Sönmez    Özgür Emre Yıldırım     1
Name: title, Length: 48186, dtype: int64
```

From the above we can see that Director **"Rajiv Chilaka"** is the most popular director in netflix and with the cast members **"Julie Tejwani"** and **"Rajesh Kava"**, he directed about 19 films with each one of them and they got streamed in the platform.

==> We are having the date_added feature is of type object, we can convert it to the date-time type feature to extract further insights from it.

In [43]:
```
exp_df['date_added'] = pd.to_datetime(exp_df['date_added'])
exp_df.head()
```

Out[43]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90 min	Documentaries
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Dramas
3	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Mysteries
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows

Types of content available in different countries

In [44]:
```
country_cont = exp_df.groupby('country')['type'].unique().to_frame().reset_index()
country_cont
```

Out[44]:

	country	type
0		[TV Show, Movie]
1	Afghanistan	[Movie]
2	Albania	[Movie]
3	Algeria	[Movie]
4	Angola	[Movie]
...	...	...
118	Vatican City	[Movie]
119	Venezuela	[Movie]
120	Vietnam	[Movie]
121	West Germany	[TV Show, Movie]
122	Zimbabwe	[Movie]

123 rows × 2 columns

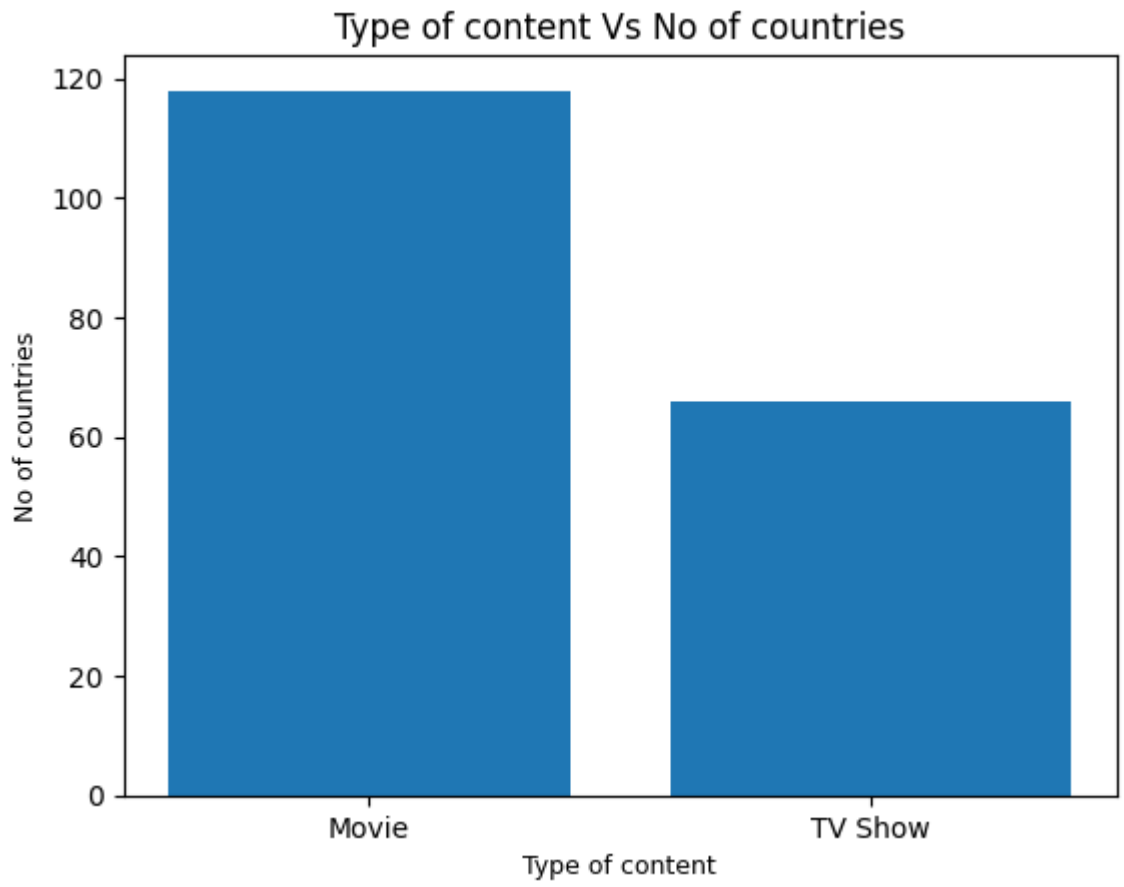
From the above we can see that many of the country consumes only Movie's and only few countries have the audience who consume both Movies and TV Shows

In [45]:
```
type_con = country_cont.explode('type').groupby('type').agg(func="count").reset_index().rename({'country':'Total no of
type_con
```

Out[45]:

	type	Total no of countries
0	Movie	118
1	TV Show	66

In [46]:
```
plt.title('Type of content Vs No of countries')
plt.bar(type_con['type'],type_con['Total no of countries'])
plt.xlabel('Type of content',fontsize=9)
plt.ylabel('No of countries',fontsize=9)
plt.show()
```



There are **118** countries in total that consumes Movie type content and there are **66** countries that consume TV show type contents.

In [47]:
```
onlyMovie = country_cont.loc[country_cont['type'].isin([["Movie"]])].explode('type')
onlyMovie
```

Out[47]:

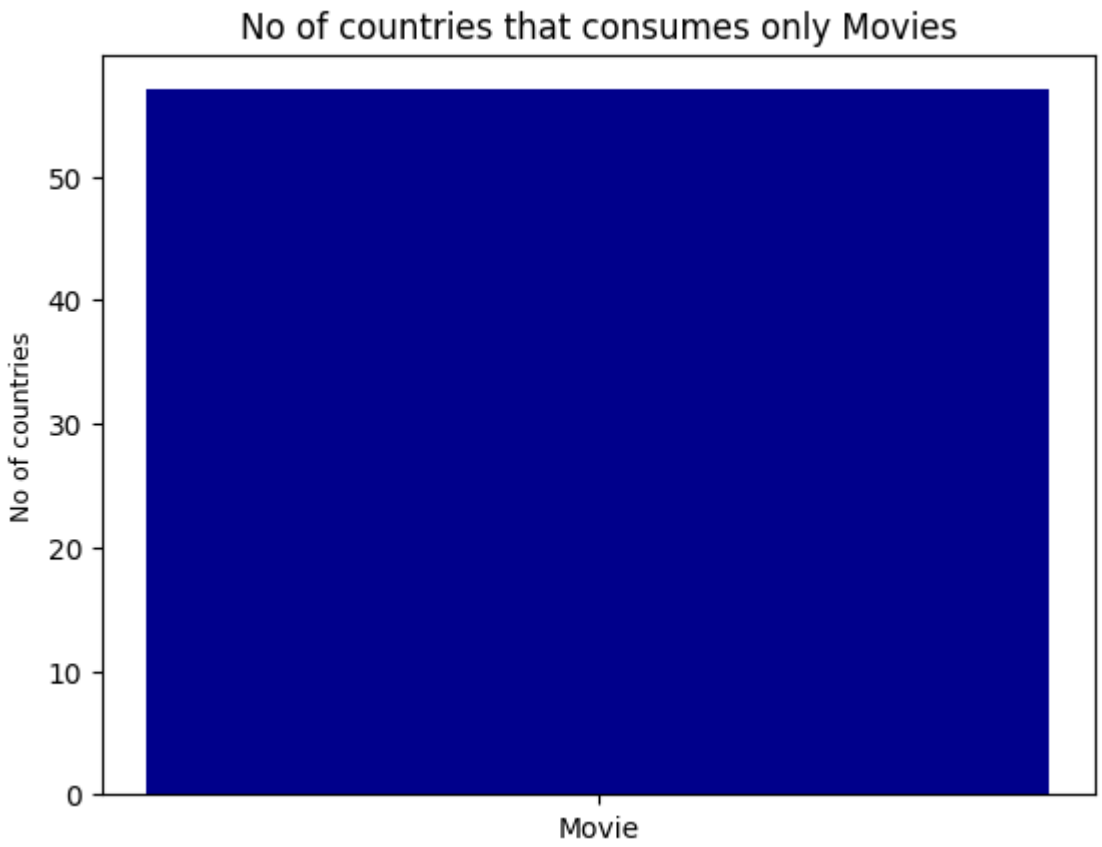
	country	type
1	Afghanistan	Movie
2	Albania	Movie
3	Algeria	Movie
4	Angola	Movie
6	Armenia	Movie
10	Bahamas	Movie
11	Bangladesh	Movie
14	Bermuda	Movie
15	Botswana	Movie
17	Bulgaria	Movie
18	Burkina Faso	Movie
19	Cambodia	Movie
20	Cameroon	Movie
22	Cayman Islands	Movie
31	Dominican Republic	Movie
32	East Germany	Movie
33	Ecuador	Movie
35	Ethiopia	Movie
38	Georgia	Movie
40	Ghana	Movie
42	Guatemala	Movie
48	Iran	Movie
49	Iraq	Movie
53	Jamaica	Movie
56	Kazakhstan	Movie
57	Kenya	Movie
59	Latvia	Movie
61	Liechtenstein	Movie
62	Lithuania	Movie
64	Malawi	Movie
69	Mongolia	Movie
70	Montenegro	Movie
71	Morocco	Movie
72	Mozambique	Movie
73	Namibia	Movie
74	Nepal	Movie
77	Nicaragua	Movie
81	Palestine	Movie
82	Panama	Movie
83	Paraguay	Movie
84	Peru	Movie
87	Portugal	Movie
89	Qatar	Movie
90	Romania	Movie
92	Samoa	Movie
95	Serbia	Movie
97	Slovakia	Movie
98	Slovenia	Movie
99	Somalia	Movie
102	Soviet Union	Movie
104	Sri Lanka	Movie
105	Sudan	Movie
112	Uganda	Movie
118	Vatican City	Movie

	country	type
119	Venezuela	Movie
120	Vietnam	Movie

In [48]:
```
onlyMovie['country'].nunique()
```

Out[48]: 57

In [49]:
```
plt.title("No of countries that consumes only Movies")
plt.bar(onlyMovie['type'].unique(),onlyMovie['type'].value_counts().values,color="darkblue")
plt.ylabel("No of countries",fontsize=9)
plt.show()
```



There are **57** Countries in the world that consumes Movie or there are **57** countries where only movies are streamed in the netflix platform.

In [50]:
```
only_tvshow = country_cont.loc[country_cont['type'].isin([["TV Show"]])].explode('type')
only_tvshow
```

Out[50]:

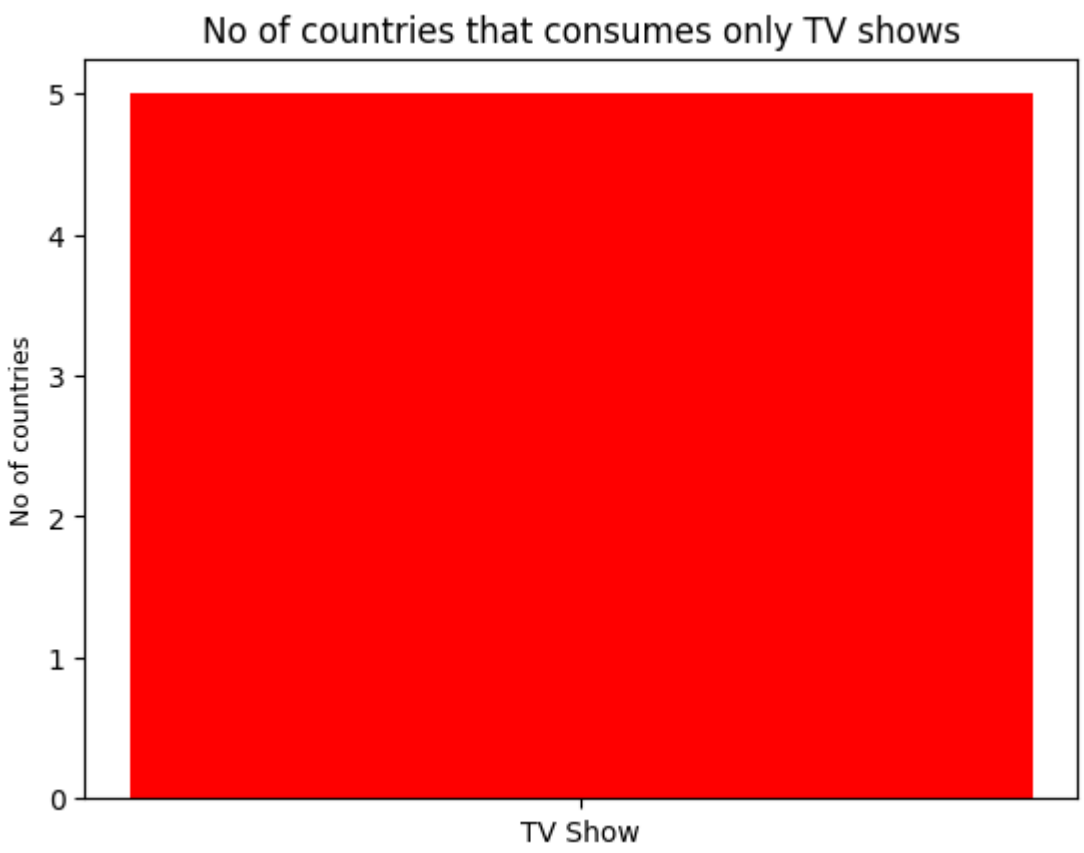
	country	type
9	Azerbaijan	TV Show
12	Belarus	TV Show
27	Cuba	TV Show
28	Cyprus	TV Show
88	Puerto Rico	TV Show

In [51]:
```
onlyTv = only_tvshow.groupby('type')['country'].count().to_frame().reset_index()
onlyTv
```

Out[51]:

	type	country
0	TV Show	5

In [52]:
```
plt.title("No of countries that consumes only TV shows")
plt.bar(onlyTv['type'],onlyTv['country'],color="red")
plt.ylabel('No of countries',fontsize=9)
plt.show()
```



There are **5** countries in the world where the people there consumes only TV shows in netflix platform (or) there are **5** countries where netflix streams only TV shows and not movies.

In [53]:
```
both_content = country_cont.loc[~(country_cont['type'].isin([["TV Show"]])) & ~(country_cont['type'].isin([["Movie"]])
both_content
```

Out[53]:

	country	type
0		[TV Show, Movie]
5	Argentina	[Movie, TV Show]
7	Australia	[TV Show, Movie]
8	Austria	[Movie, TV Show]
13	Belgium	[TV Show, Movie]
...	...	...
114	United Arab Emirates	[Movie, TV Show]
115	United Kingdom	[Movie, TV Show]
116	United States	[Movie, TV Show]
117	Uruguay	[TV Show, Movie]
121	West Germany	[TV Show, Movie]

61 rows × 2 columns

In [54]:
```
both_content['type'] = both_content['type'].apply(lambda x :"Both streamed")
both_content
```

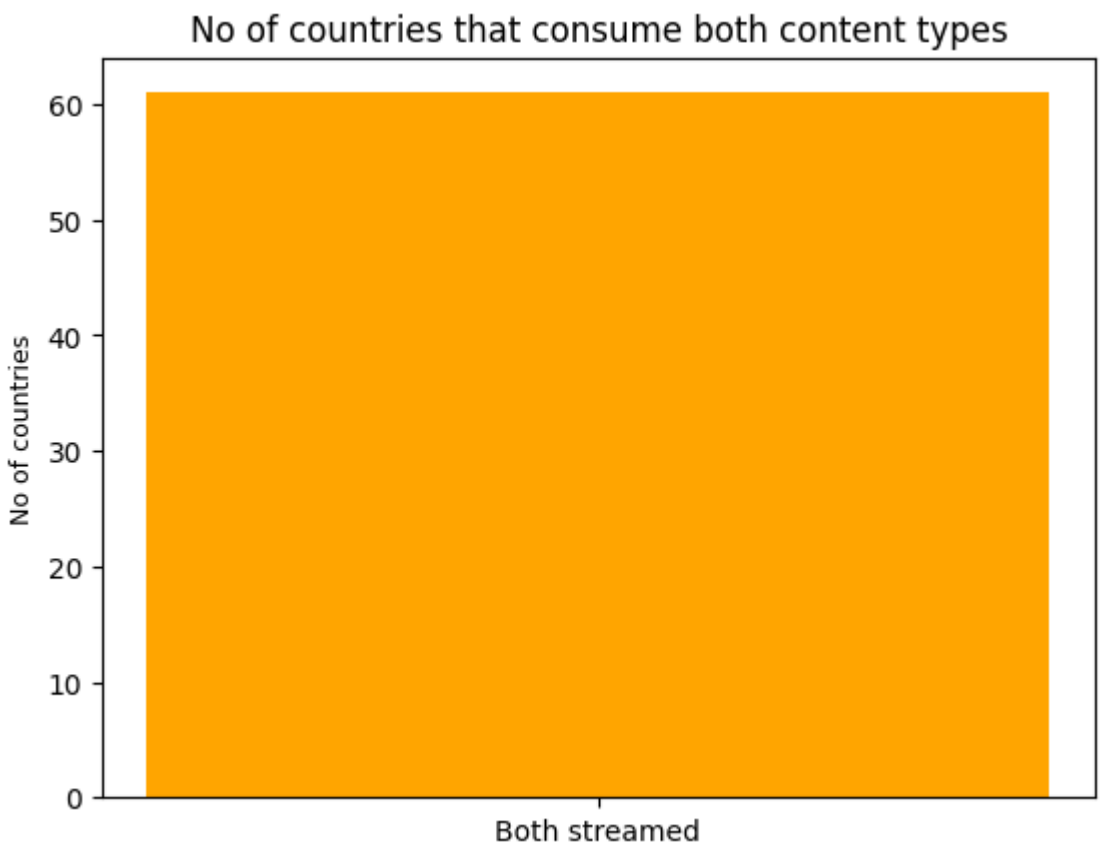
```
<ipython-input-54-672283e61b5d>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  both_content['type'] = both_content['type'].apply(lambda x :"Both streamed")
```

Out[54]:

	country	type
0		Both streamed
5	Argentina	Both streamed
7	Australia	Both streamed
8	Austria	Both streamed
13	Belgium	Both streamed
...	...	...
114	United Arab Emirates	Both streamed
115	United Kingdom	Both streamed
116	United States	Both streamed
117	Uruguay	Both streamed
121	West Germany	Both streamed

61 rows × 2 columns

In [55]:
```
plt.title("No of countries that consume both content types")
plt.bar(both_content['type'].unique(),both_content['type'].value_counts(),color="orange")
plt.ylabel("No of countries",fontsize=9)
plt.show()
```



There are 62 Countries where both Movies and TV shows are streamed and consumed by people through netflix.

Country wise Number of Movies and TV Shows

In [56]:
```
exp_df.head()
```

Out[56]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90 min	Documentaries
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Dramas
3	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Mysteries
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows

In [57]:
```
types = exp_df.groupby(['country','type'])['title'].nunique().to_frame().reset_index()
types
```

Out[57]:

	country	type	title
0		Movie	1
1		TV Show	1
2	Afghanistan	Movie	1
3	Albania	Movie	1
4	Algeria	Movie	3
...	...	...	...
179	Venezuela	Movie	4
180	Vietnam	Movie	7
181	West Germany	Movie	3
182	West Germany	TV Show	2
183	Zimbabwe	Movie	3

184 rows × 3 columns

Pivotting to get more clarity from the data

In [58]:
```
types = types.pivot(index='country',columns='type',values='title').reset_index()
types.columns.name=None
types
```

Out[58]:

	country	Movie	TV Show
0		1.0	1.0
1	Afghanistan	1.0	NaN
2	Albania	1.0	NaN
3	Algeria	3.0	NaN
4	Angola	1.0	NaN
...	...	...	...
118	Vatican City	1.0	NaN
119	Venezuela	4.0	NaN
120	Vietnam	7.0	NaN
121	West Germany	3.0	2.0
122	Zimbabwe	3.0	NaN

123 rows × 3 columns

For each country from above we can see the Number of movies and Number of TV Shows that got released

In [59]:
```
temp = types.loc[types['country']=='India']
temp
```

Out[59]:

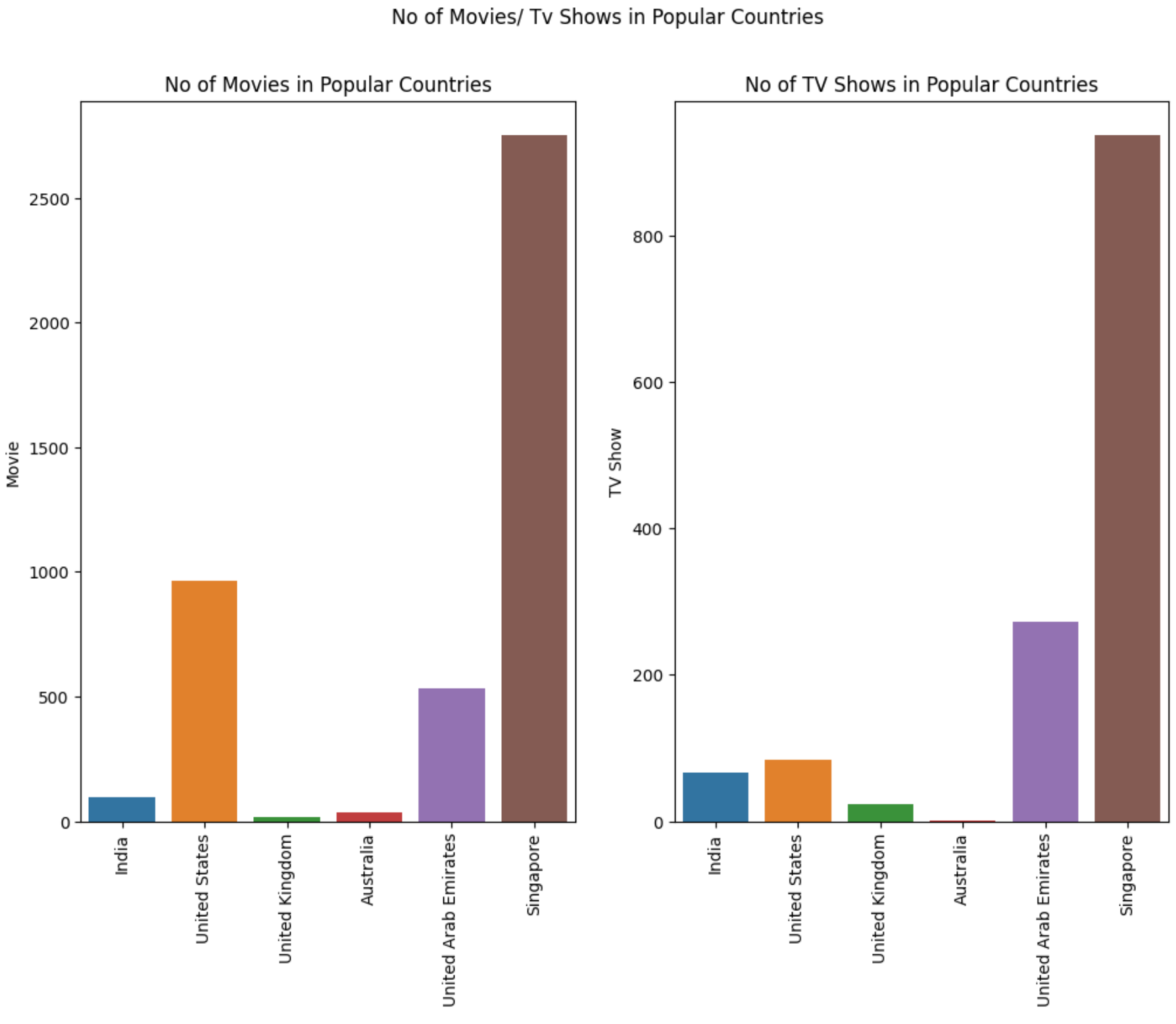
	country	Movie	TV Show
46	India	962.0	84.0

In India, We can see that the number of movies and TV Shows added in Netflix is **962** and **84** respectively.

In [60]:
```
major_countries = ['India','United States','United Kingdom','Australia','United Arab Emirates','Singapore']  #Popular
```

Movie and TV Show distribution in Popular countries:

In [61]:
```
plt.figure(figsize=(12,8))
plt.subplot(1,2,1)
plt.xticks(rotation=90)
plt.title("No of Movies in Popular Countries")
sns.barplot(x=major_countries,y='Movie',data=types.loc[types['country'].isin(major_countries)])
plt.subplot(1,2,2)
plt.xticks(rotation=90)
plt.title("No of TV Shows in Popular Countries")
sns.barplot(x=major_countries,y='TV Show',data=types.loc[types['country'].isin(major_countries)])
plt.suptitle("No of Movies/ Tv Shows in Popular Countries")
plt.show()
```



Number of Movies released over the years!

In [62]:
```
exp_df.head()
```

Out[62]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90 min	Documentaries
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Dramas
3	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Mysteries
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows

The dataset comprises of both Movie and TV show related data. we have to separate out them into separate dataframe/datasets.

In [63]:
```
movies = exp_df.loc[exp_df['type']=="Movie"]
movies.reset_index(drop=True,inplace=True)
movies
```

Out[63]:

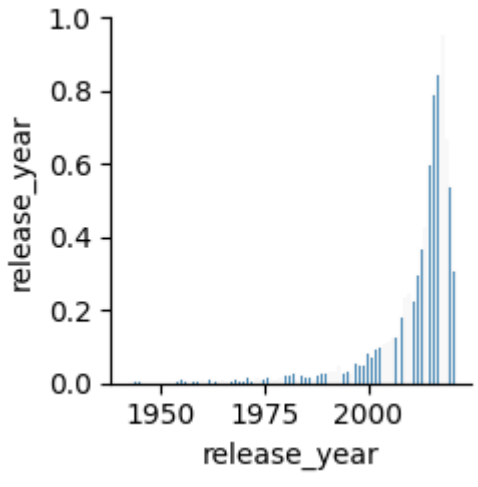
	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90 min	Documentaries
1	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91 min	Children & Family Movies
2	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91 min	Children & Family Movies
3	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	NaN	2021-09-24	2021	PG	91 min	Children & Family Movies
4	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Kimiko Glenn	NaN	2021-09-24	2021	PG	91 min	Children & Family Movies
...	...	...	...	...	...	...	...	...	...	...	...
145838	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2019-03-02	2015	TV-14	111 min	International Movies
145839	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2019-03-02	2015	TV-14	111 min	Music & Musicals
145840	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	111 min	Dramas
145841	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	111 min	International Movies
145842	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	111 min	Music & Musicals

145843 rows × 11 columns

Pairplot to visualize the distribution and trends they follow between different numerical features of the movies dataframe

In [64]:
```
plt.figure(figsize=(12,8))
sns.pairplot(movies)
```

Out[64]:
```
<seaborn.axisgrid.PairGrid at 0x7f46f0a38bb0>

<Figure size 1200x800 with 0 Axes>
```



From above, by noticing the added_year and duration features, we can see that the duration of movies have been increased significantly over the years. There is an increasing trend that can be noticed.

The number of unique release years present in the movies dataframe

In [65]:
```
movies['release_year'].nunique()
```

Out[65]:
```
73
```

The dataframe comprises of movies that were released over the last 73 years!

In [66]:
```
movies['release_year'].unique()
```

Out[66]:
```
array([2020, 2021, 1993, 1996, 1998, 1997, 2010, 2013, 2017, 1975, 1978,
       1983, 1987, 2012, 2001, 2002, 2003, 2004, 2011, 2008, 2009, 2007,
       2005, 2006, 2018, 2019, 1994, 2015, 1982, 1989, 2014, 1990, 1991,
       1999, 2016, 1986, 1984, 1980, 1961, 2000, 1995, 1985, 1992, 1976,
       1959, 1988, 1981, 1972, 1964, 1954, 1979, 1958, 1956, 1963, 1970,
       1973, 1960, 1974, 1966, 1971, 1962, 1969, 1977, 1967, 1968, 1965,
       1945, 1946, 1942, 1955, 1944, 1947, 1943])
```

Number of movies released each year

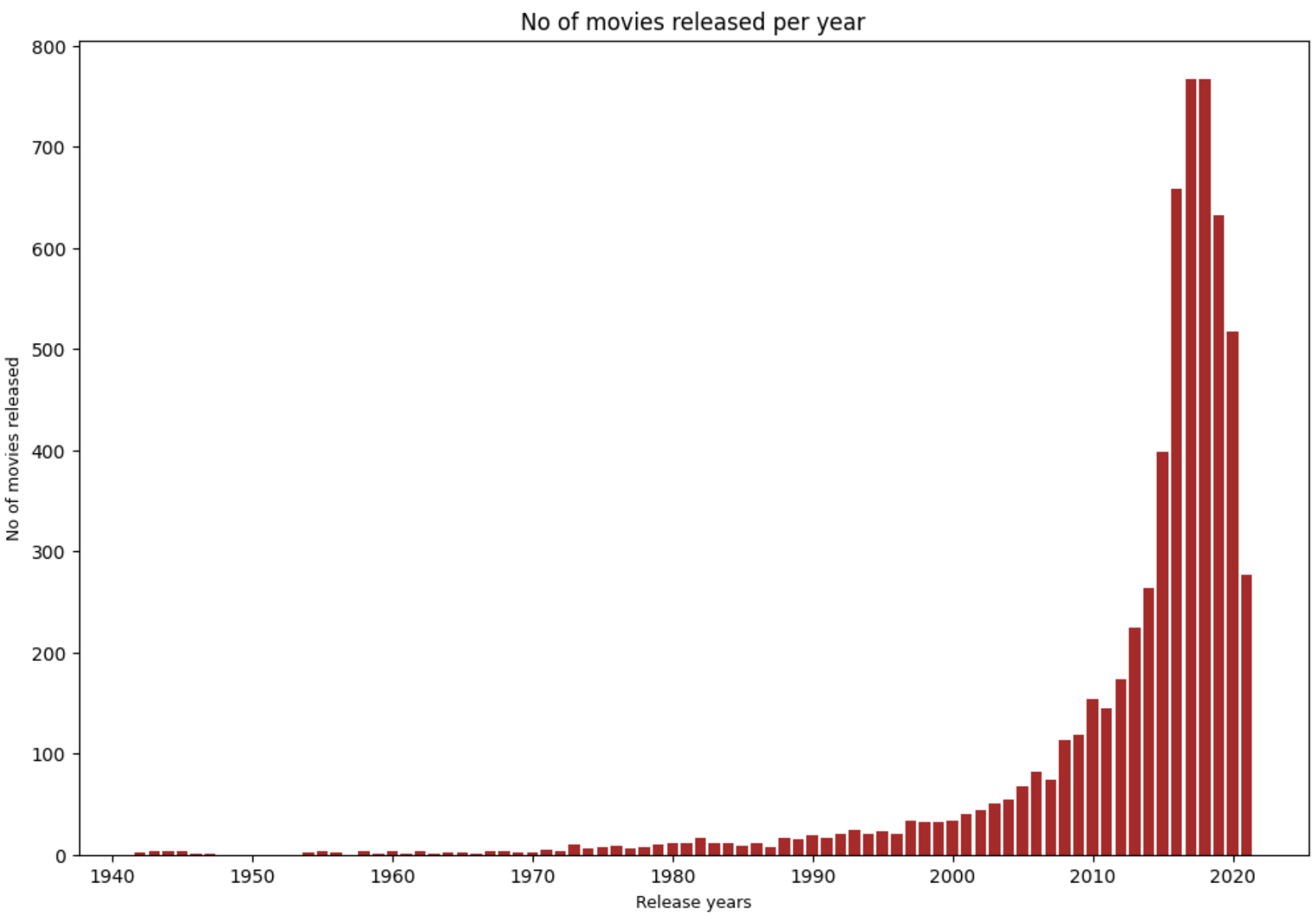
In [67]:
```
movie_count = movies.groupby('release_year')['title'].nunique().to_frame(name="Number of movies released").reset_index
movie_count
```

Out[67]:

	release_year	Number of movies released
0	1942	2
1	1943	3
2	1944	3
3	1945	3
4	1946	1
...	...	...
68	2017	767
69	2018	767
70	2019	633
71	2020	517
72	2021	277

73 rows × 2 columns

In [68]:
```
plt.figure(figsize=(12,8))
plt.title("No of movies released per year")
plt.bar(movie_count['release_year'],movie_count['Number of movies released'],color="brown")
plt.xlabel("Release years",fontsize=9)
plt.ylabel("No of movies released",fontsize=9)
plt.show()
```



In 1940's and 1950's, we can see that there were only afew movies released since it was the initial stages of entertainment industry. Video camera's was just invented during this phase and only short clippings that were shot were filmed for people.

Over the last 20 to 30 years, ie from 1980-1990's to 2020, there has been a huge growth in the entertainment industry as result of numerous technological advancements and developments. The raise of OTT platforms is also one of the reason to this surge in the trend enabling the people to get exposed to different kinds of content from different countries.

In [69]:
```
movie_count.loc[movie_count['release_year']==1990]
```

Out[69]:

	release_year	Number of movies released
41	1990	19

The number of movies released in the year 1990 was just 19.

In [70]:
```
movie_count.loc[movie_count['release_year']==2017]
```

Out[70]:

	release_year	Number of movies released
68	2017	767

The Number of movies that were released in the year 2017 was about 767. There was about **3940%** increase in the number of movies released per year.

Comparison of TV Shows Vs Movies

In [71]:
```
TVShow = exp_df.loc[exp_df['type']=="TV Show"]
TVShow.reset_index(drop=True,inplace=True)
TVShow
```

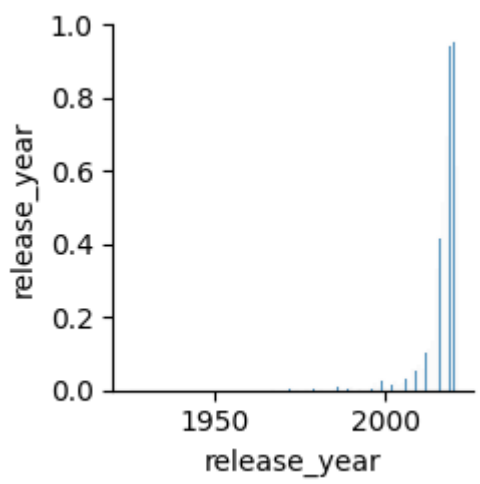
Out[71]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Dramas
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Mysteries
3	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Dramas
...	...	...	...	...	...	...	...	...	...	...	...
56143	s8801	TV Show	Zindagi Gulzar Hai	NaN	Hina Khawaja Bayat	Pakistan	2016-12-15	2012	TV-PG	1 Season	Romantic TV Shows
56144	s8801	TV Show	Zindagi Gulzar Hai	NaN	Hina Khawaja Bayat	Pakistan	2016-12-15	2012	TV-PG	1 Season	TV Dramas
56145	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2018	TV-Y7	2 Seasons	Kids' TV
56146	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2018	TV-Y7	2 Seasons	Korean TV Shows
56147	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2018	TV-Y7	2 Seasons	TV Comedies

56148 rows × 11 columns

In [72]:
```
sns.pairplot(TVShow)
```

Out[72]:
```
<seaborn.axisgrid.PairGrid at 0x7f46ee3da830>
```



From the above pairplot, we can notice the trends between different numerical features of the TV Show dataframe and by observing here, we could infer that there has been a significant increase in the duration of a TV show over the years.

In [73]:
```
TVShow['release_year'].nunique()
```

Out[73]:
```
46
```

Dataset have the data related to the TV shows that were streamed for last 46 years!

In [74]:
```
TVShow['release_year'].unique()
```

Out[74]:
```
array([2021, 2020, 2018, 2014, 1994, 2015, 2013, 2019, 2017, 2016, 2012,
       1992, 2002, 2009, 2011, 2005, 2008, 2010, 2007, 2001, 2006, 1993,
       1997, 2003, 1945, 1999, 1998, 2000, 2004, 1986, 1995, 1925, 1972,
       1974, 1988, 1991, 1977, 1979, 1990, 1996, 1981, 1946, 1985, 1967,
       1989, 1963])
```

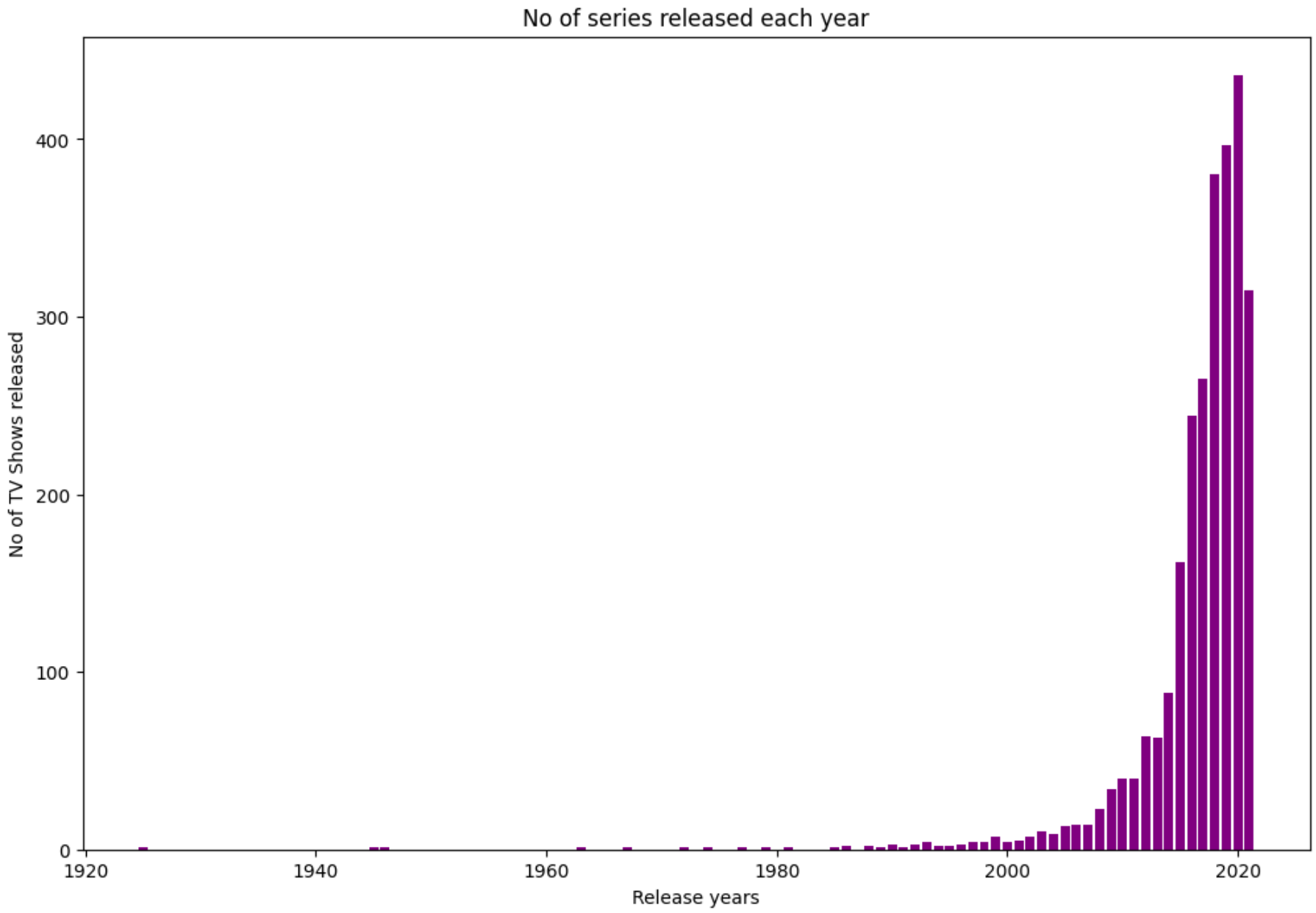
In [75]:
```
TVShow_count = TVShow.groupby('release_year')['title'].nunique().to_frame(name="No of series released").reset_index()
TVShow_count.head(10)
```

Out[75]:

	release_year	No of series released
0	1925	1
1	1945	1
2	1946	1
3	1963	1
4	1967	1
5	1972	1
6	1974	1
7	1977	1
8	1979	1
9	1981	1

From the above we can see there has been a significant increase in the number of TV shows released over the years.

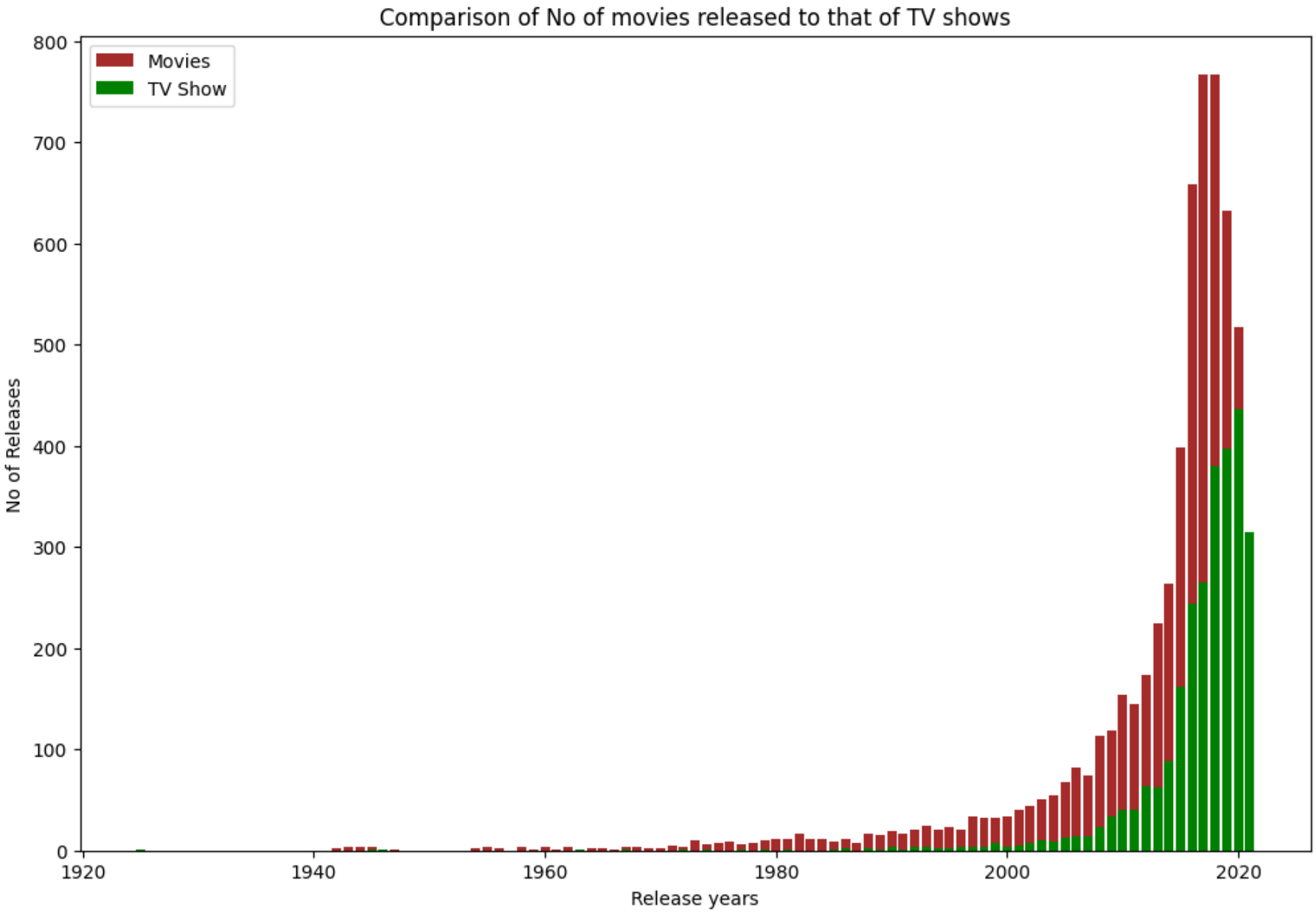
In [76]:
```
plt.figure(figsize=(12,8))
plt.bar(TVShow_count['release_year'],TVShow_count['No of series released'],color="purple")
plt.title("No of series released each year")
plt.xlabel("Release years")
plt.ylabel("No of TV Shows released")
plt.show()
```



From the above plot we can see that untill the early 2000's, there were only a few TV shows got released. People were more used to Movies that time and thus producers were not ready to produce TV shows.

After 2010, the emergence of different OTT platforms enabled people to consume more number of TV shows both from Native and Other countries. We can see that Most number of TV shows got released in the year 2020.

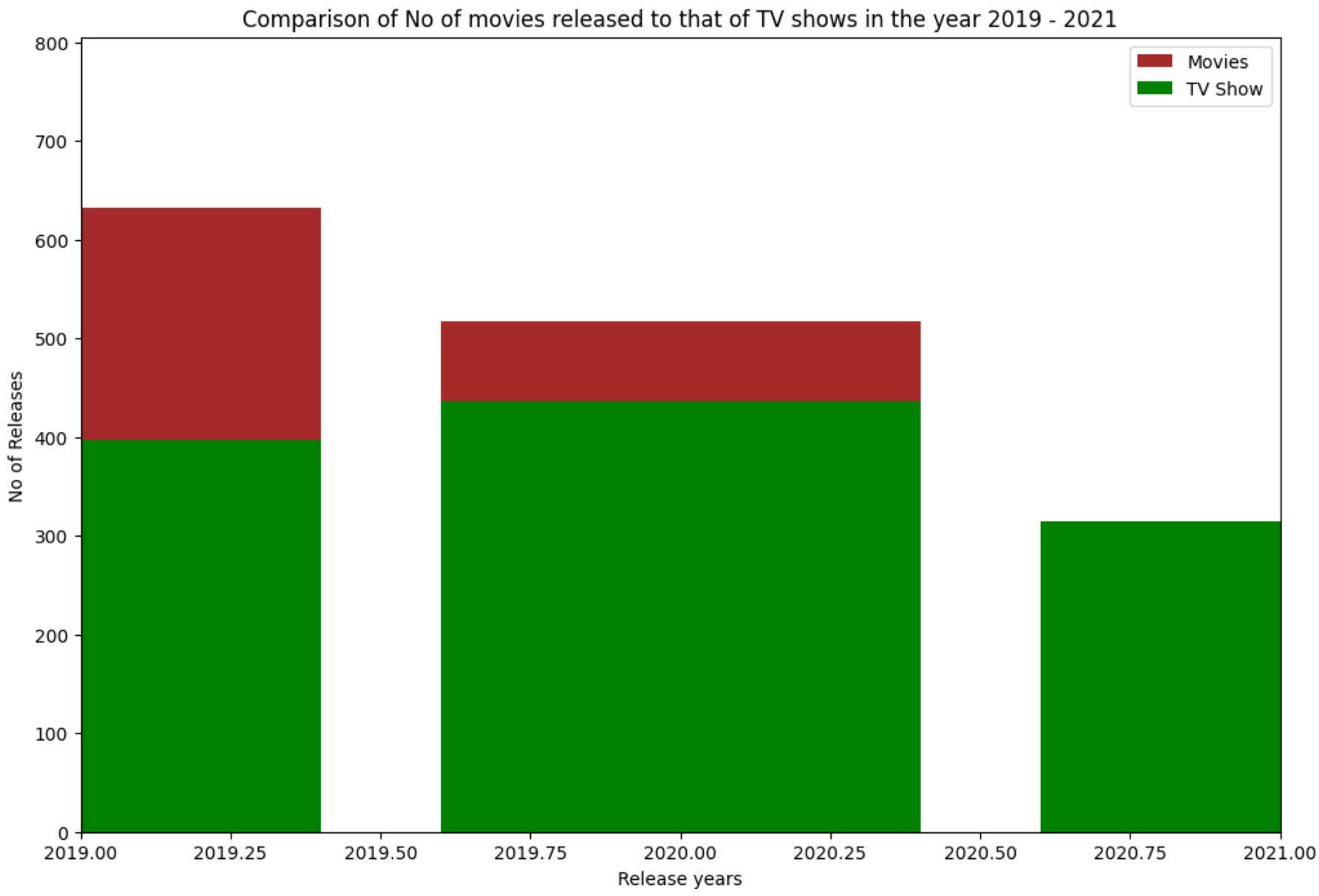
In [77]:
```
plt.figure(figsize=(12,8))
plt.bar(movie_count['release_year'],movie_count['Number of movies released'],color="brown",label="Movies")
plt.bar(TVShow_count['release_year'],TVShow_count['No of series released'],color="green",label="TV Show")
plt.xlabel("Release years")
plt.ylabel("No of Releases")
plt.legend()
plt.title("Comparison of No of movies released to that of TV shows")
plt.show()
```



Clearly from the above we can see that there were many number of movie releases over the years when compared to that of the TV Show releases.

From 2019 to 2021

In [78]:
```
plt.figure(figsize=(12,8))
plt.bar(movie_count['release_year'],movie_count['Number of movies released'],color="brown",label="Movies")
plt.bar(TVShow_count['release_year'],TVShow_count['No of series released'],color="green",label="TV Show")
plt.xlabel("Release years")
plt.ylabel("No of Releases")
plt.legend()
plt.title("Comparison of No of movies released to that of TV shows in the year 2019 - 2021")
plt.xlim(left=2019,right=2021)
plt.show()
```



In the recent years ie from 2019 to 2021, we can observe that the ratio of number of movies to TV shows had been reduced significantly indicating that Netflix had started to focus more on TV shows than Movies.

Movies and TV Shows

Movies:

We will convert the duration feature of object type to int type by extracting the numerical details alone from them.

There are few missing values in this feature, those missing values can be imputed later.

In [79]:
```
movies['duration'] = movies['duration'].apply(lambda x:x if type(x)==float else int(x.split(' ')[0]))
```

```
<ipython-input-79-ed95f4a9df8a>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  movies['duration'] = movies['duration'].apply(lambda x:x if type(x)==float else int(x.split(' ')[0]))
```

In [80]:
```
movies.head()
```

Out[80]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90.0	Documentaries
1	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
2	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
3	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
4	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies

The average run time of the movies is:

In [81]:
```
mruntime = movies.groupby(['title'])['duration'].unique().to_frame().reset_index().explode('duration')
mruntime
```

Out[81]:

	title	duration
0	#Alive	99.0
1	#AnneFrank - Parallel Stories	95.0
2	#FriendButMarried	102.0
3	#FriendButMarried 2	104.0
4	#Roxy	105.0
...	...	...
6126	Maj Rati Keteki	117.0
6127	Mayurakshi	100.0
6128	Kuch Bheege Alfaaz	110.0
6129	반드시 잡는다	110.0
6130	최강전사 미니특공대 : 영웅의 탄생	68.0

6131 rows × 2 columns

In [82]:
```
mruntime['duration'].mean()
```

Out[82]:
```
99.57718668407311
```

The average runtime of movies that got released across globally is **99.57** Minutes.

Average run-time of movies based on the country, where it was released.

In [83]:
```
loc_avgtime = movies.groupby('country')['duration'].unique().to_frame().reset_index()
loc_avgtime['duration'] = loc_avgtime['duration'].apply(np.mean)
loc_avgtime
```

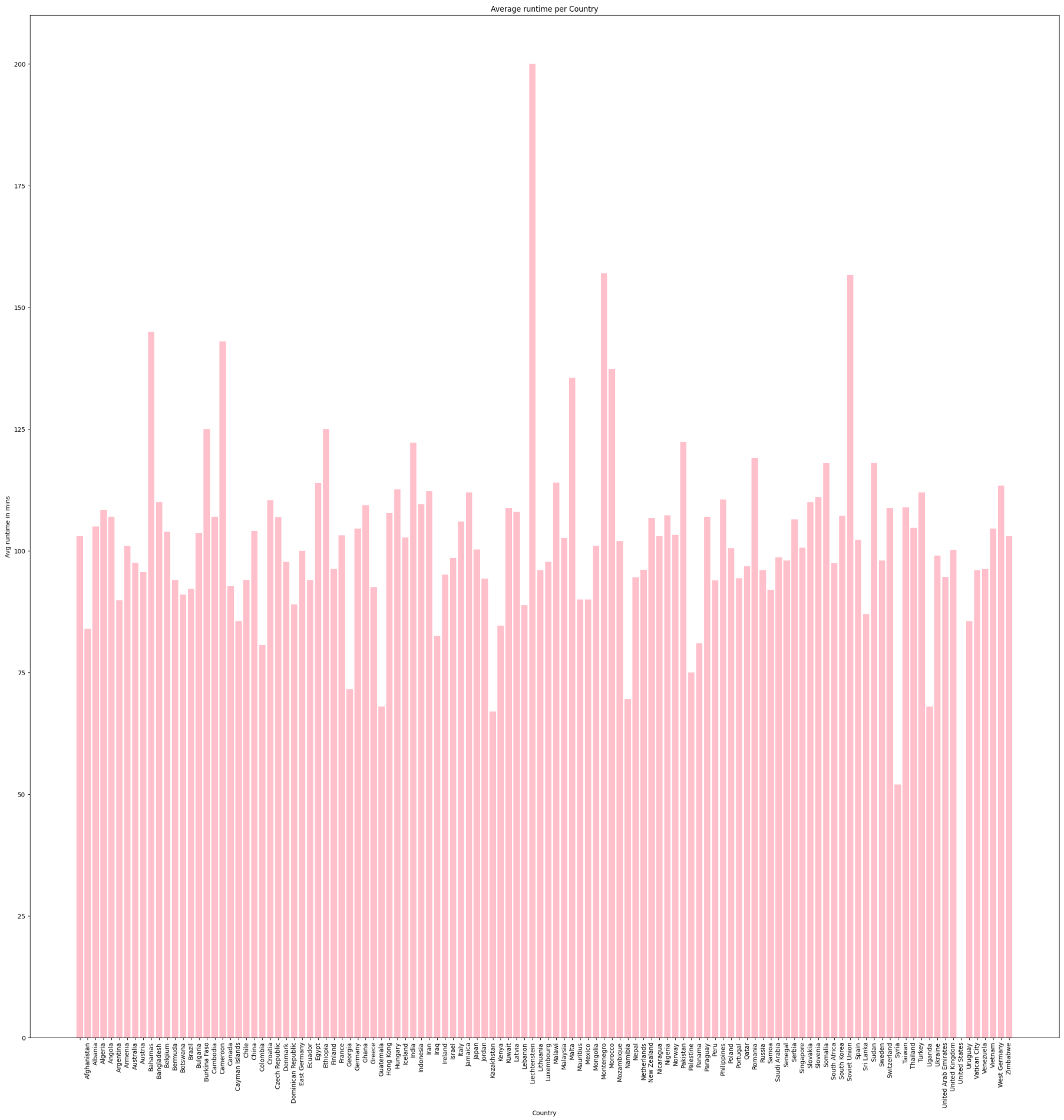
Out[83]:

	country	duration
0		103.000000
1	Afghanistan	84.000000
2	Albania	105.000000
3	Algeria	108.333333
4	Angola	107.000000
...	...	...
113	Vatican City	96.000000
114	Venezuela	96.250000
115	Vietnam	104.500000
116	West Germany	113.333333
117	Zimbabwe	103.000000

118 rows × 2 columns

From the above we can see the avg runtime of movies across different countries.

In [84]:
```
plt.figure(figsize=(30,30))
plt.bar(loc_avgtime['country'],loc_avgtime['duration'],color="pink")
plt.title("Average runtime per Country")
plt.xticks(rotation=90)
plt.xlabel("Country")
plt.ylabel("Avg runtime in mins")
plt.show()
```



Duration of the most number of movies

In [85]:
```
movies.head()
```

Out[85]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90.0	Documentaries
1	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
2	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
3	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
4	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies

In [86]:
```
movies['duration'].unique()
```

Out[86]:
```
array([ 90.,  91., 125., 104., 127.,  67.,  94., 161.,  61., 166., 147.,
       103.,  97., 106., 111., 110., 105.,  96., 124., 116.,  98.,  23.,
       115., 122.,  99.,  88., 100., 102.,  93.,  95.,  85.,  83., 113.,
        13., 182.,  48., 145.,  87.,  92.,  80., 117., 128., 119., 143.,
       114., 118., 108.,  63., 121., 142., 154., 120.,  82., 109., 101.,
        86., 229.,  76.,  89., 156., 112., 107., 129., 135., 136., 165.,
       150., 133.,  70.,  84., 140.,  78.,  64.,  59., 139.,  69., 148.,
       189., 141., 130., 138.,  81., 132., 123.,  65.,  68.,  66.,  62.,
        74., 131.,  39.,  46.,  38., 126., 155., 159., 137.,  12., 273.,
        36.,  34.,  77.,  60.,  49.,  58.,  72., 204., 212.,  25.,  73.,
        29.,  47.,  32.,  35.,  71., 149.,  33.,  15.,  54., 224., 162.,
        37.,  75.,  79.,  55., 158., 164., 173., 181., 185.,  21.,  24.,
        51., 151.,  42.,  22., 134., 177.,  52.,  14.,  53.,   8.,  57.,
        28.,  50.,   9.,  26.,  45., 171.,  27.,  44., 146.,  20., 157.,
        17., 203.,  41.,  30., 194., 233., 237., 230., 195., 253., 152.,
       190., 160., 208., 180., 144.,   5., 174., 170., 192., 209., 187.,
       172.,  16., 186.,  11., 193., 176.,  56., 169.,  40.,  10.,   3.,
       168., 312., 153., 214.,  31., 163.,  19.,  nan, 179.,  43., 200.,
       196., 167., 178., 228.,  18., 205., 201., 191.])
```

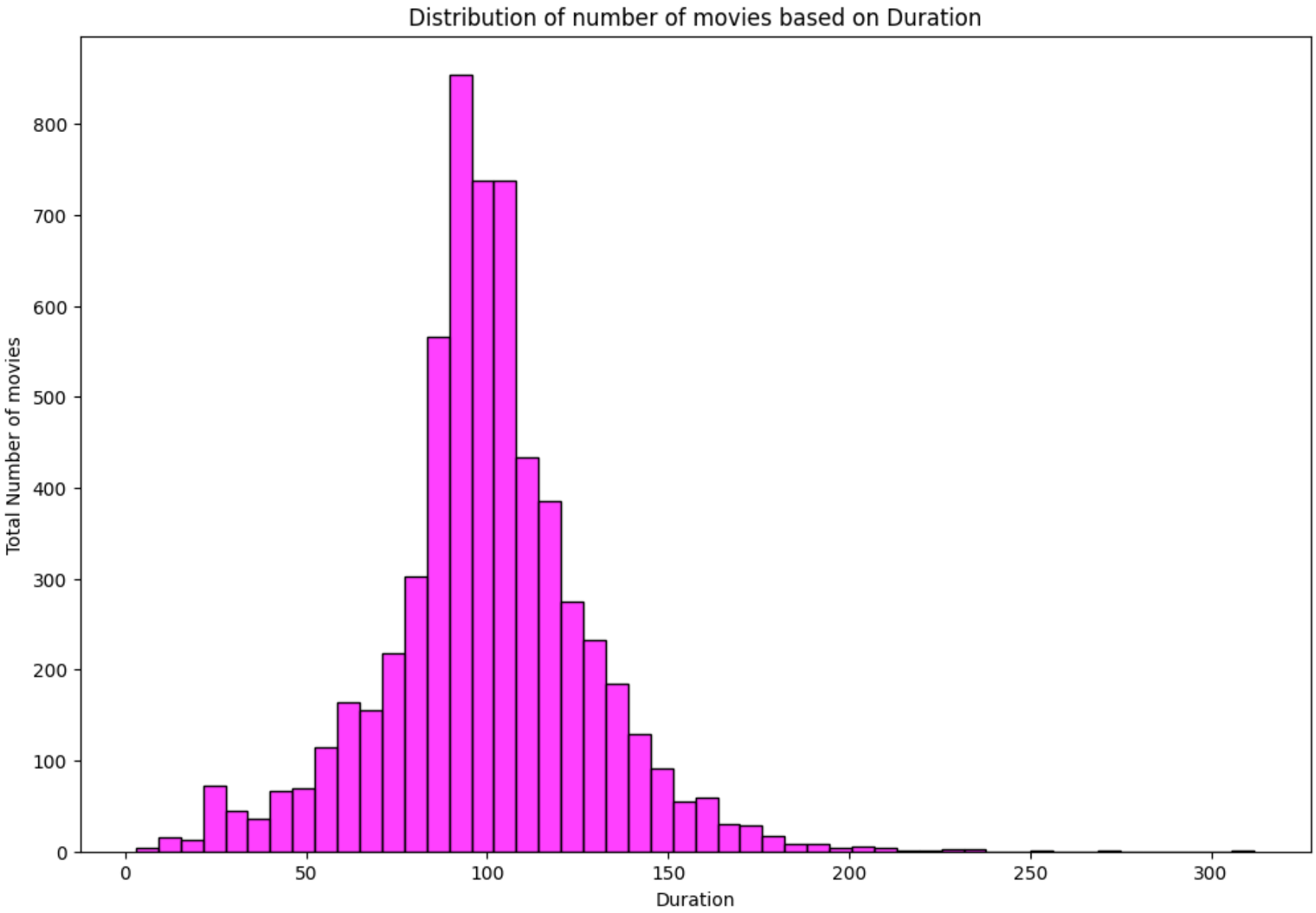
In [87]:
```
mv_duration = movies.groupby(['title'])['duration'].unique().to_frame().reset_index()
mv_duration['duration'] = mv_duration['duration'].apply(lambda x:x.mean() if type(x) != float else x)
mv_duration
```

Out[87]:

	title	duration
0	#Alive	99.0
1	#AnneFrank - Parallel Stories	95.0
2	#FriendButMarried	102.0
3	#FriendButMarried 2	104.0
4	#Roxy	105.0
...	...	...
6126	Maj Rati Keteki	117.0
6127	Mayurakshi	100.0
6128	Kuch Bheege Alfaaz	110.0
6129	반드시 잡는다	110.0
6130	최강전사 미니특공대 : 영웅의 탄생	68.0

6131 rows × 2 columns

In [88]:
```
plt.figure(figsize=(12,8))
sns.histplot(data=mv_duration,x='duration',bins=50,color="magenta")
plt.title("Distribution of number of movies based on Duration")
plt.ylabel("Total Number of movies")
plt.xlabel("Duration")
plt.show()
```



From the above we can conclude that most number of movies that are added in Netflix have a run-time of **85 to 110 minutes** approximately.

Genres that are popular in each of the country.

In [89]:
```
movies.head()
```

Out[89]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90.0	Documentaries
1	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
2	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
3	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
4	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies

We can groupby the dataset by Country and Genre to get the most popular Genre for each country.

In [90]:
```
mv_genre = movies.groupby(['country','listed_in'])['title'].agg(func="nunique").to_frame().reset_index()
mv_genres_max = mv_genre.groupby(['country'])['title'].agg(func="max").to_frame().reset_index()
```

In [91]:
```
mv_genre
```

Out[91]:

	country	listed_in	title
0		Dramas	1
1		Independent Movies	1
2		International Movies	1
3	Afghanistan	Documentaries	1
4	Afghanistan	International Movies	1
...	...	...	...
889	West Germany	Thrillers	1
890	Zimbabwe	Comedies	1
891	Zimbabwe	Documentaries	2
892	Zimbabwe	International Movies	3
893	Zimbabwe	Romantic Movies	1

894 rows × 3 columns

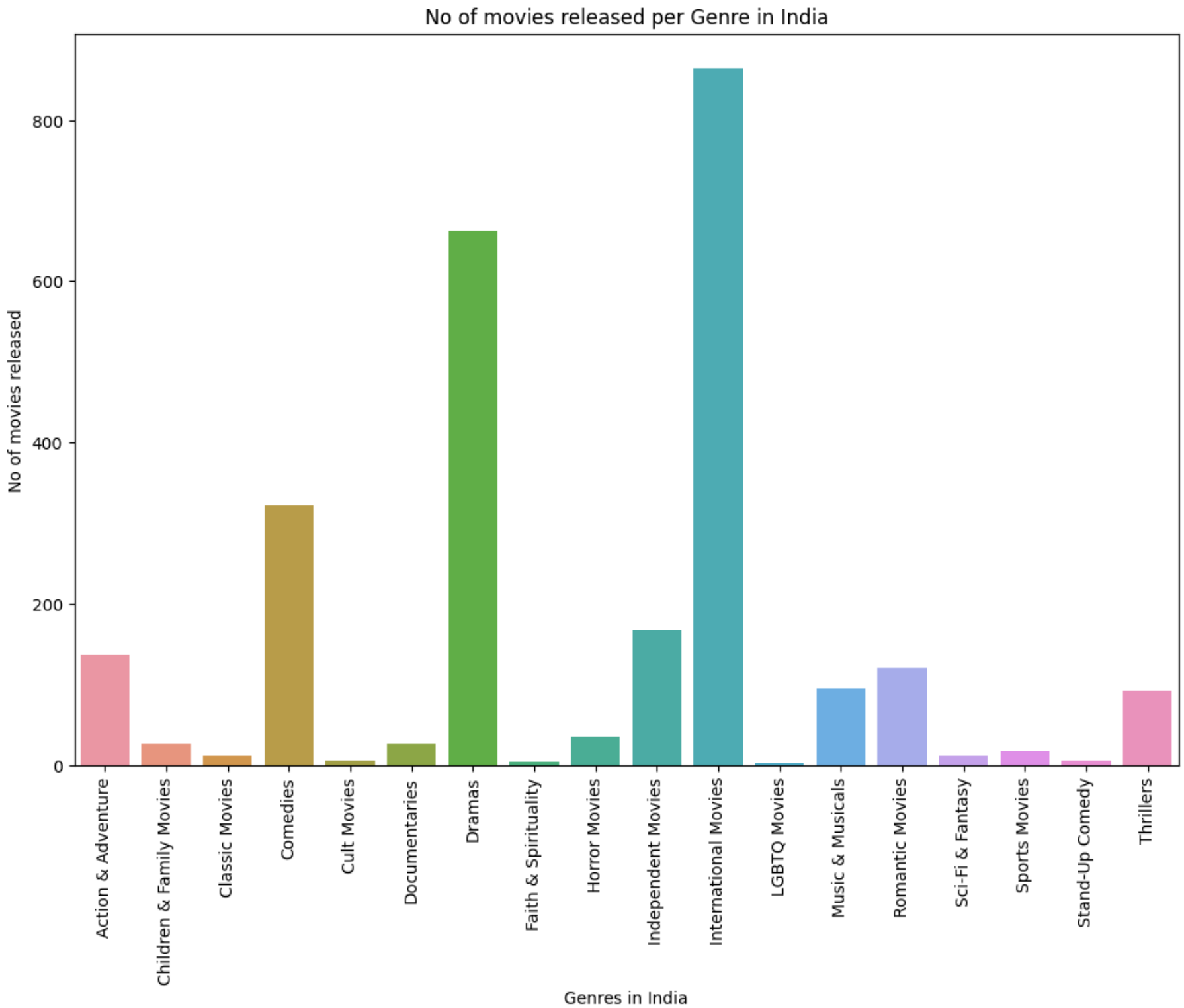
The number of movies released per Genre in India

In [92]:
```
mv_india = mv_genre.loc[mv_genre['country']=='India']
mv_india
```

Out[92]:

	country	listed_in	title
319	India	Action & Adventure	137
320	India	Children & Family Movies	26
321	India	Classic Movies	11
322	India	Comedies	323
323	India	Cult Movies	5
324	India	Documentaries	27
325	India	Dramas	662
326	India	Faith & Spirituality	4
327	India	Horror Movies	35
328	India	Independent Movies	167
329	India	International Movies	864
330	India	LGBTQ Movies	2
331	India	Music & Musicals	96
332	India	Romantic Movies	120
333	India	Sci-Fi & Fantasy	12
334	India	Sports Movies	17
335	India	Stand-Up Comedy	6
336	India	Thrillers	92

In [93]:
```
plt.figure(figsize=(12,8))
sns.barplot(data=mv_india,x='listed_in',y='title')
plt.xticks(rotation=90)
plt.xlabel("Genres in India")
plt.ylabel("No of movies released")
plt.title("No of movies released per Genre in India")
plt.show()
```



In [94]:
```
mv_genres = mv_genre.merge(mv_genres_max,on=["country","title"],how="inner")
mv_genres
```

Out[94]:

	country	listed_in	title
0		Dramas	1
1		Independent Movies	1
2		International Movies	1
3	Afghanistan	Documentaries	1
4	Afghanistan	International Movies	1
...	...	...	...
169	Venezuela	Documentaries	3
170	Venezuela	International Movies	3
171	Vietnam	International Movies	7
172	West Germany	International Movies	2
173	Zimbabwe	International Movies	3

174 rows × 3 columns

Thus from the above, we get the popular genres in each of the country with the number of movies released under that Genre in that country.

Popular movie actor in each of the country

In [95]:
```
movies.head()
```

Out[95]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90.0	Documentaries
1	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
2	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
3	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
4	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies

In [96]:
```
pop_actor = movies.groupby(['country','cast'])['title'].nunique().to_frame().reset_index()
pop_actor
```

Out[96]:

	country	cast	title
0	Areen Omari		1
1	Khaled Abol El Naga		1
2	Maisa Abd Elhadi		1
3	Malak Ermileh		1
4	Nisreen Faour		1
...	...	...	...
37993	Zimbabwe	Kudzai Sevenzo	1
37994	Zimbabwe	Michael Kudakwashe	1
37995	Zimbabwe	Tendai Nguni	1
37996	Zimbabwe	Tendaiishe Chitima	1
37997	Zimbabwe	Zihlo	1

37998 rows × 3 columns

In [97]:
```
pop = pop_actor.groupby(['country'])['title'].agg(func="max").to_frame().reset_index()
pop_actor = pop_actor.merge(pop,on=['country','title'],how='inner')
pop_actor.sort_values(by='title',ascending=False,inplace=True)
pop_actor.reset_index(inplace=True,drop=True)
pop_actor
```

Out[97]:

	country	cast	title
0	India	Anupam Kher	40
1	United States	Samuel L. Jackson	21
2	Turkey	Demet Akbağ	13
3	Egypt	Ahmed Helmy	13
4	Egypt	Hassan Hosny	13
...	...	...	...
991	Finland	Kaija Pakarinen	1
992	Finland	Laura Birn	1
993	Finland	Lauri Tanskanen	1
994	Finland	Leslie Jones	1
995	Zimbabwe	Zihlo	1

996 rows × 3 columns

From the above we can see the most popular actor in each of the country and the number of their films that got released in the Netflix platform.

For example: In India, the Most popular actor is Anupam Kher and about **40** of his films got released on the netflix platform.

Best Time to launch a movie

What is the best month to launch a movie ?

In [98]:
```
movies.head()
```

Out[98]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90.0	Documentaries
1	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
2	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
3	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies
4	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies

In [99]:
```
movies['added_month'] = movies['date_added'].dt.month_name()
movies['added_year'] = movies['date_added'].dt.year
movies.head()
```

```
<ipython-input-99-9f7bf90983c0>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  movies['added_month'] = movies['date_added'].dt.month_name()
<ipython-input-99-9f7bf90983c0>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  movies['added_year'] = movies['date_added'].dt.year
```

Out[99]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	added_month	added_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90.0	Documentaries	September	2021
1	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies	September	2021
2	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies	September	2021
3	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies	September	2021
4	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies	September	2021

In [100]:
```
mv_months = movies.groupby(['added_year','added_month'])['title'].nunique().to_frame().reset_index()
mv_month = mv_months.groupby('added_year')['title'].agg(func="max").to_frame().reset_index()
mv_month = mv_months.merge(mv_month,on=['added_year','title'],how='inner')
mv_month
```

Out[100]:

	added_year	added_month	title
0	2008	January	1
1	2009	May	1
2	2009	November	1
3	2010	November	1
4	2011	October	11
5	2012	December	1
6	2012	February	1
7	2012	November	1
8	2013	December	2
9	2013	November	2
10	2014	December	5
11	2015	December	14
12	2016	December	52
13	2017	October	97
14	2018	October	146
15	2019	November	187
16	2020	January	152
17	2021	July	169

Over the years, we can see that more number of Movies were added in the netflix platform in the Months of October, November and December. Since it is a festive season and people tend to go on vacation in this period, It would be the best time to launch new movies in the platform.

I would recommend netflix to release more number of movies in this festive season.

Popular Director in different Genres

In [101]:
```
popgen_direc = movies.groupby(['listed_in','director'])['title'].nunique().to_frame().reset_index()
popgen = popgen_direc.groupby('listed_in')['title'].agg(func="max").to_frame().reset_index()
popgen_direc = popgen_direc.merge(popgen,on=['listed_in','title'],how="inner")
popgen_direc
```

Out[101]:

	listed_in	director	title
0	Action & Adventure	Don Michael Paul	9
1	Anime Features	Toshiya Shinohara	7
2	Children & Family Movies	Rajiv Chilaka	22
3	Classic Movies	Youssef Chahine	8
4	Comedies	David Dhawan	9
5	Cult Movies	Mike Clattenburg	3
6	Documentaries	Vlad Yudin	6
7	Dramas	Youssef Chahine	12
8	Faith & Spirituality	David Batty	5
9	Horror Movies	Rocky Soraya	6
10	Independent Movies	Noah Baumbach	5
11	Independent Movies	Paul Thomas Anderson	5
12	International Movies	Cathy Garcia-Molina	13
13	LGBTQ Movies	Jun Lana	2
14	LGBTQ Movies	Leigh Janiak	2
15	LGBTQ Movies	Matt Kugelman	2
16	LGBTQ Movies	Saratswadee Wongsomphet	2
17	Movies	Louis C.K.	3
18	Music & Musicals	Matt Askem	6
19	Romantic Movies	Cathy Garcia-Molina	8
20	Sci-Fi & Fantasy	Lana Wachowski	4
21	Sci-Fi & Fantasy	Lilly Wachowski	4
22	Sports Movies	Vlad Yudin	4
23	Stand-Up Comedy	Jan Suter	21
24	Thrillers	David Fincher	4
25	Thrillers	Rathindran R Prasad	4

In [102]:
```
gen_direc = popgen_direc.groupby('listed_in')['director'].unique().to_frame().reset_index()
gen_direc
```

Out[102]:

	listed_in	director
0	Action & Adventure	[Don Michael Paul]
1	Anime Features	[Toshiya Shinohara]
2	Children & Family Movies	[Rajiv Chilaka]
3	Classic Movies	[Youssef Chahine]
4	Comedies	[David Dhawan]
5	Cult Movies	[Mike Clattenburg]
6	Documentaries	[Vlad Yudin]
7	Dramas	[Youssef Chahine]
8	Faith & Spirituality	[David Batty]
9	Horror Movies	[Rocky Soraya]
10	Independent Movies	[Noah Baumbach, Paul Thomas Anderson]
11	International Movies	[Cathy Garcia-Molina]
12	LGBTQ Movies	[Jun Lana, Leigh Janiak, Matt Kugelman, Sarats...
13	Movies	[Louis C.K.]
14	Music & Musicals	[Matt Askem]
15	Romantic Movies	[Cathy Garcia-Molina]
16	Sci-Fi & Fantasy	[Lana Wachowski, Lilly Wachowski]
17	Sports Movies	[Vlad Yudin]
18	Stand-Up Comedy	[Jan Suter]
19	Thrillers	[David Fincher, Rathindran R Prasad]

From the above we get the list of most popular directors for each genre.

Number of movies in each of the rating

In [103]:
```
movies.head()
```

Out[103]:

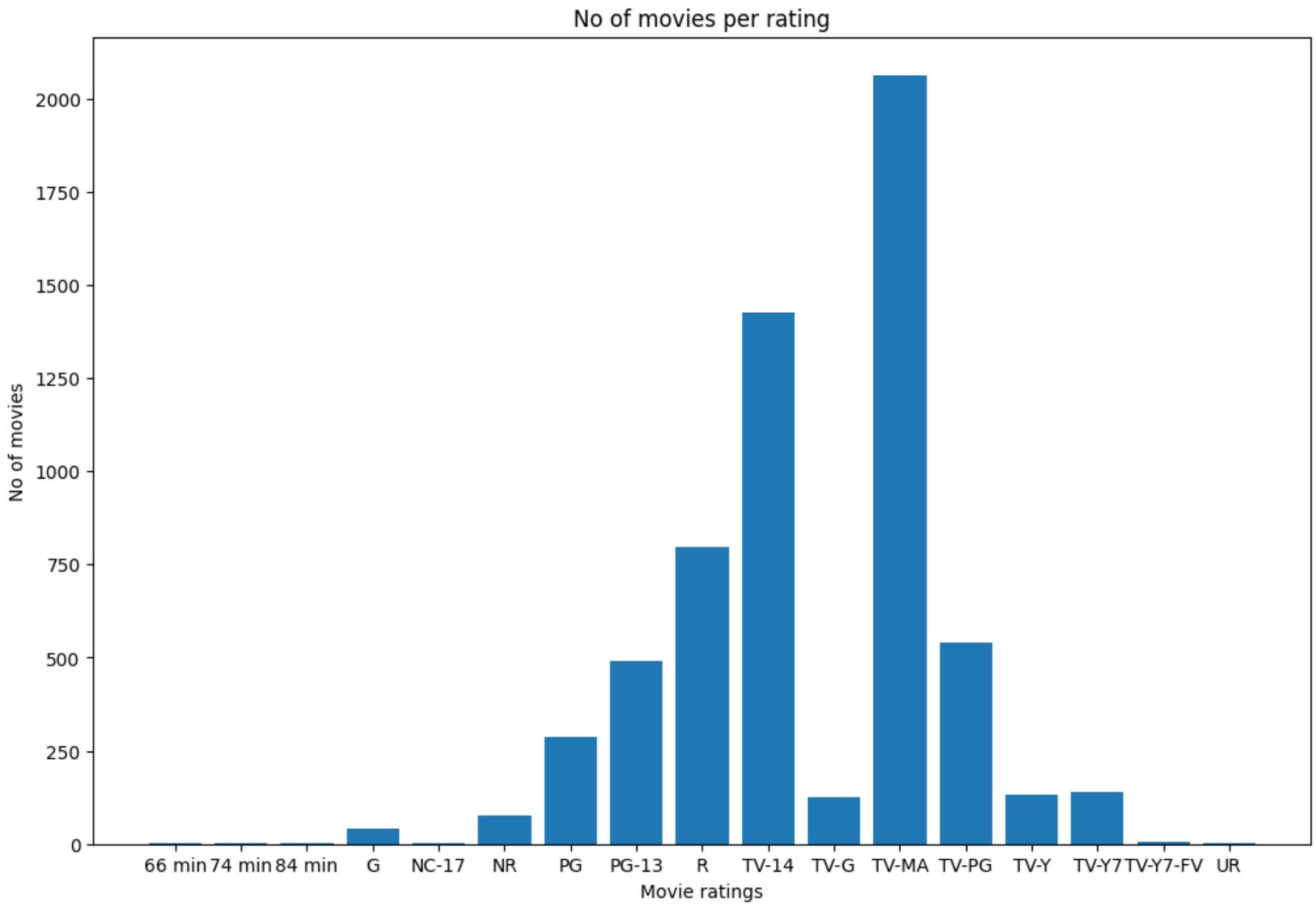
	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	added_month	added_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	90.0	Documentaries	September	2021
1	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies	September	2021
2	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Vanessa Hudgens	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies	September	2021
3	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies	September	2021
4	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Kimiko Glenn	NaN	2021-09-24	2021	PG	91.0	Children & Family Movies	September	2021

In [104]:
```
mv_rating = movies.groupby(['rating'])['title'].nunique().to_frame().reset_index()
mv_rating
```

Out[104]:

	rating	title
0	66 min	1
1	74 min	1
2	84 min	1
3	G	41
4	NC-17	3
5	NR	75
6	PG	287
7	PG-13	490
8	R	797
9	TV-14	1427
10	TV-G	126
11	TV-MA	2062
12	TV-PG	540
13	TV-Y	131
14	TV-Y7	139
15	TV-Y7-FV	5
16	UR	3

In [105]:
```
plt.figure(figsize=(12,8))
plt.bar(x=mv_rating['rating'],height=mv_rating['title'])
plt.xlabel("Movie ratings")
plt.ylabel("No of movies")
plt.title("No of movies per rating")
plt.show()
```



Thus from the above, we can see that most number of movies that were released in the netflix platform is of TV_MA rating, which means most of the movies have that level of explicit sexuality, graphic violence and/or strong language along with mature themes making them unsuitable for children 16 or younger.

TV Shows:

For TV shows the run-time is mentioned in Seasons, thus we have to split them accordingly and continue the processing.

In [106]:
```
TVShow.isna().sum()
```

Out[106]:
```
show_id            0
type               0
title              0
director       49358
cast             818
country         5698
date_added       158
release_year       0
rating            58
duration           0
listed_in          0
dtype: int64
```

There are no missing values in the duration feature of TVShow dataset.

In [107]:
```
TVShow['duration'] = TVShow['duration'].apply(lambda x:int(x.split(' ')[0]))
TVShow.rename({"duration":"duration in seasons"},axis=1,inplace=True)
TVShow.head()
```

```
<ipython-input-107-816cf710ba81>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  TVShow['duration'] = TVShow['duration'].apply(lambda x:int(x.split(' ')[0]))
<ipython-input-107-816cf710ba81>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  TVShow.rename({"duration":"duration in seasons"},axis=1,inplace=True)
```

Out[107]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration in seasons	listed_in
0	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Mysteries
3	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas

In [108]:
```
TVruntime = TVShow.groupby('title')['duration in seasons'].unique().to_frame().reset_index().explode('duration in seas
TVruntime
```

Out[108]:

	title	duration in seasons
0	#blackAF	1
1	(Un)Well	1
2	100 Days My Prince	1
3	100 Humans	1
4	100% Hotter	1
...	...	...
2671	แผนร้ายนายเจ้าเล่ห์	1
2672	SAINT SEIYA: Knights of the Zodiac	2
2673	忍者ハットリくん	2
2674	海的儿子	1
2675	마녀사냥	1

2676 rows × 2 columns

In [109]:
```
TVruntime['duration in seasons'].mean().round()
```

Out[109]: 2.0

The TV shows that are released across globally have an average runtime of about **2 seasons**.

Average run-time of TV shows based on the country, where it was released.

In [110]:
```
tvloc_avgtime = TVShow.groupby('country')['duration in seasons'].unique().to_frame().reset_index()
tvloc_avgtime['duration in seasons'] = tvloc_avgtime['duration in seasons'].apply(np.mean)
tvloc_avgtime['duration in seasons'] = tvloc_avgtime['duration in seasons'].apply(np.round)
tvloc_avgtime
```

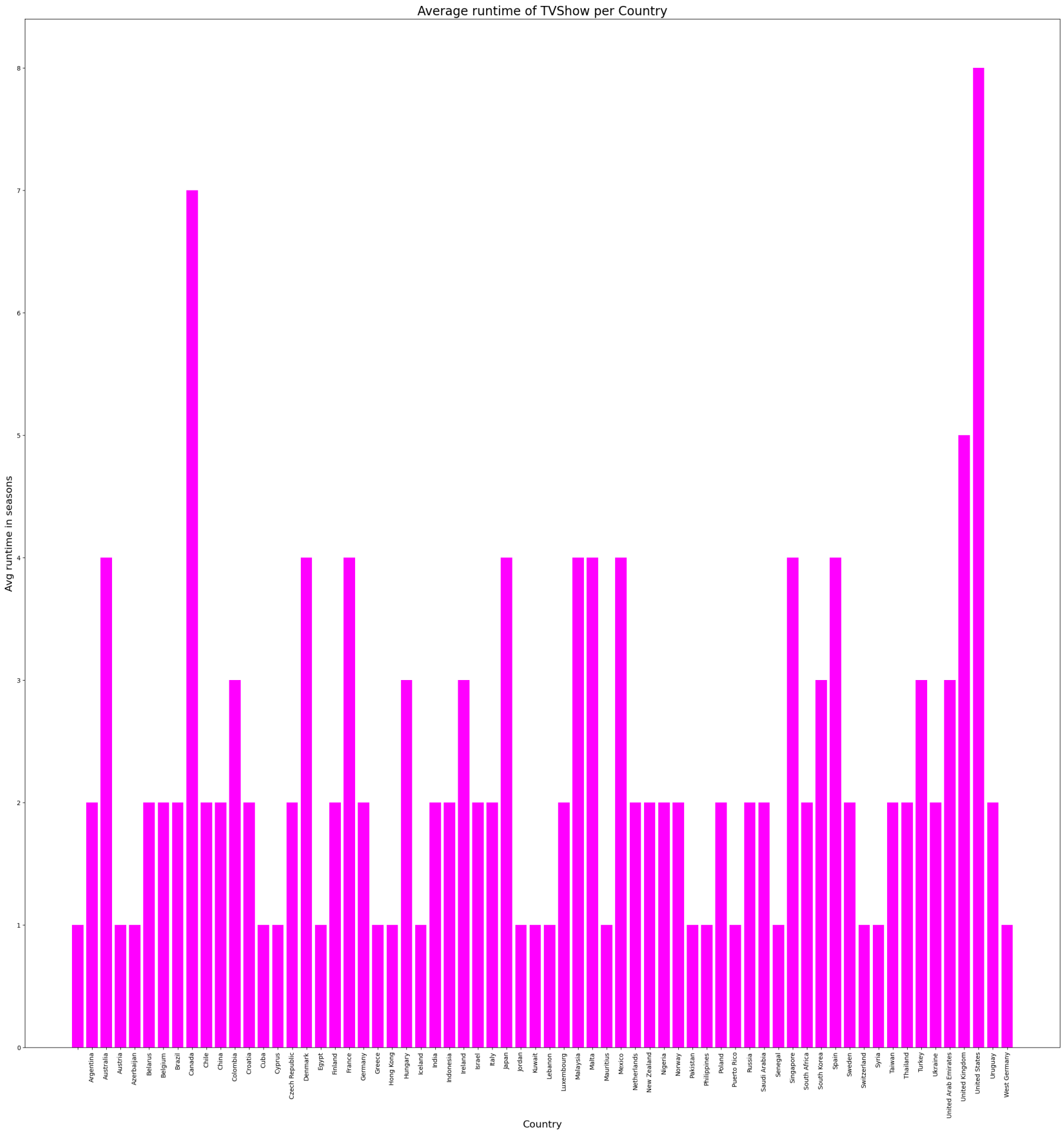
Out[110]:

	country	duration in seasons
0		1.0
1	Argentina	2.0
2	Australia	4.0
3	Austria	1.0
4	Azerbaijan	1.0
...	...	...
61	United Arab Emirates	3.0
62	United Kingdom	5.0
63	United States	8.0
64	Uruguay	2.0
65	West Germany	1.0

66 rows × 2 columns

From the above we can see the average run-time of TV shows across different countries.

In [111]:
```
plt.figure(figsize=(30,30))
plt.bar(tvloc_avgtime['country'],tvloc_avgtime['duration in seasons'],color='magenta')
plt.title("Average runtime of TVShow per Country",fontsize=20)
plt.xticks(rotation=90)
plt.xlabel("Country",fontsize=16)
plt.ylabel("Avg runtime in seasons",fontsize=16)
plt.show()
```



Duration of the Most number of TV Shows

In [112]:
```
TVShow.head()
```

Out[112]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration in seasons	listed_in
0	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Mysteries
3	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas

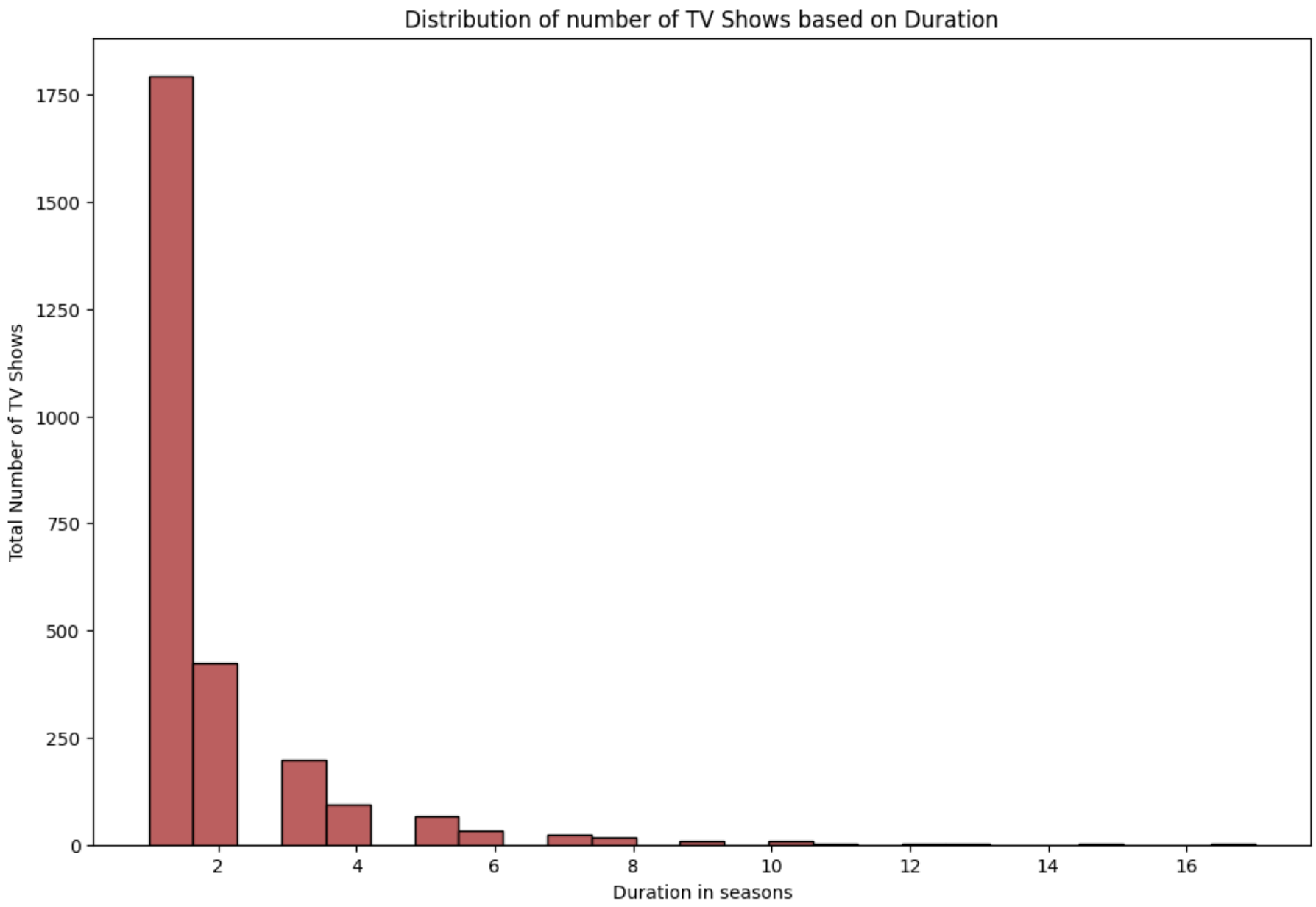
In [113]:
```
tv_duration = TVShow.groupby('title')['duration in seasons'].unique().to_frame().reset_index()
tv_duration['duration in seasons'] = tv_duration['duration in seasons'].apply(lambda x:x.mean() if type(x)!=float else
tv_duration
```

Out[113]:

	title	duration in seasons
0	#blackAF	1.0
1	(Un)Well	1.0
2	100 Days My Prince	1.0
3	100 Humans	1.0
4	100% Hotter	1.0
...	...	...
2671	แผนร้ายนายเจ้าเล่ห์	1.0
2672	SAINT SEIYA: Knights of the Zodiac	2.0
2673	忍者ハットリくん	2.0
2674	海的儿子	1.0
2675	마녀사냥	1.0

2676 rows × 2 columns

In [114]:
```
plt.figure(figsize=(12,8))
sns.histplot(data=tv_duration,x='duration in seasons',color='brown',bins=25)
plt.title("Distribution of number of TV Shows based on Duration")
plt.ylabel("Total Number of TV Shows")
plt.xlabel("Duration in seasons")
plt.show()
```



From the above we can conclude that most number of TV Shows that are added in Netflix have a run-time of **1.75 to 2 Seasons** approximately.

Genres of TV Show that are popular in each of the country.

In [115]:
```
TVShow.head()
```

Out[115]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration in seasons	listed_in
0	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Mysteries
3	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas

In [116]:
```
tv_genre = TVShow.groupby(['country','listed_in'])['title'].nunique().to_frame().reset_index()
tv_genres_max = tv_genre.groupby(['country'])['title'].agg(func="max").to_frame().reset_index()
tv_genres_max
```

Out[116]:

	country	title
0		1
1	Argentina	18
2	Australia	31
3	Austria	1
4	Azerbaijan	1
...	...	...
61	United Arab Emirates	1
62	United Kingdom	225
63	United States	258
64	Uruguay	1
65	West Germany	1

66 rows × 2 columns

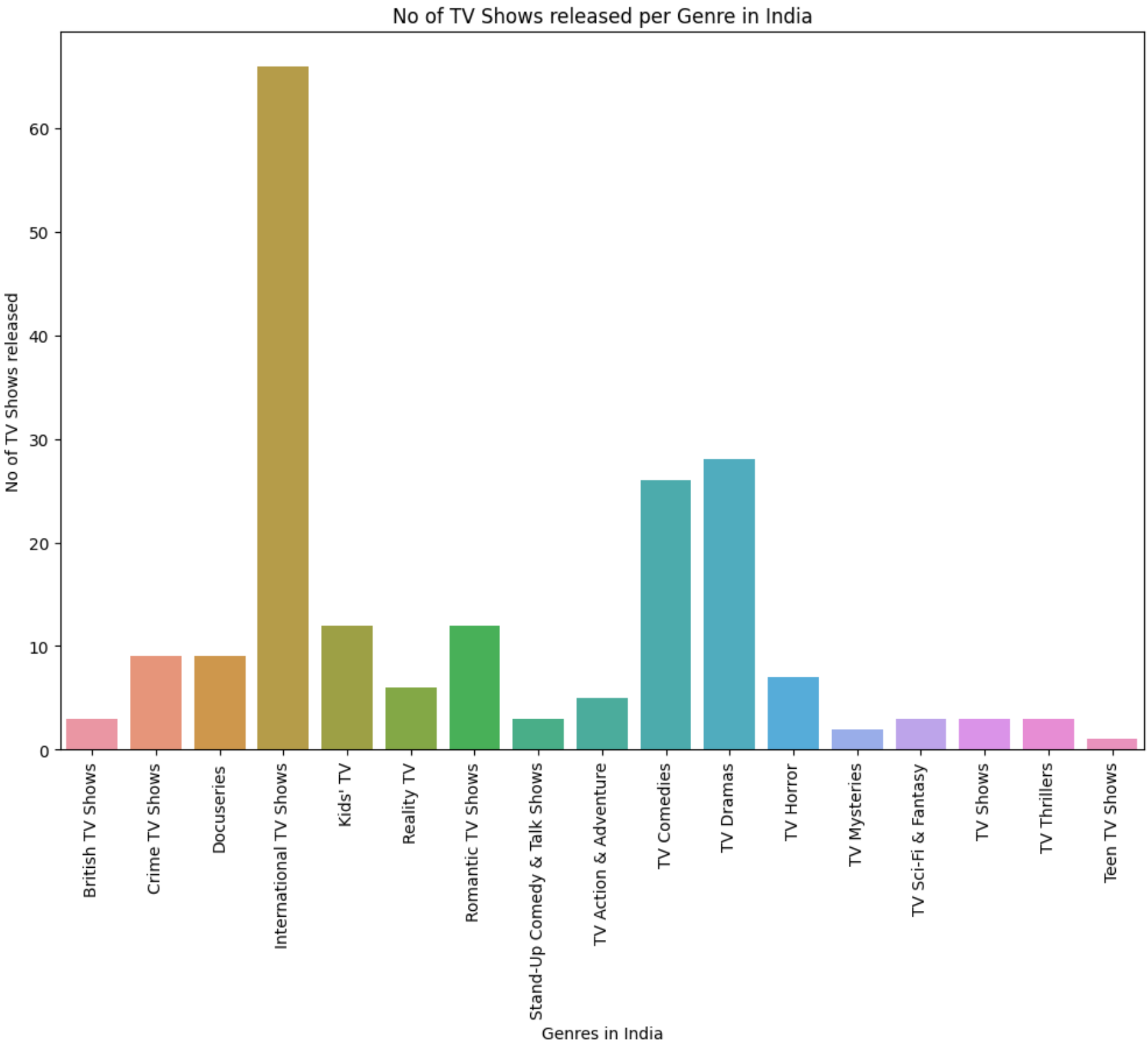
No of movies released in different Genres in India.

In [117]:
```
tv_india = tv_genre.loc[tv_genre['country']=='India']
tv_india
```

Out[117]:

	country	listed_in	title
192	India	British TV Shows	3
193	India	Crime TV Shows	9
194	India	Docuseries	9
195	India	International TV Shows	66
196	India	Kids' TV	12
197	India	Reality TV	6
198	India	Romantic TV Shows	12
199	India	Stand-Up Comedy & Talk Shows	3
200	India	TV Action & Adventure	5
201	India	TV Comedies	26
202	India	TV Dramas	28
203	India	TV Horror	7
204	India	TV Mysteries	2
205	India	TV Sci-Fi & Fantasy	3
206	India	TV Shows	3
207	India	TV Thrillers	3
208	India	Teen TV Shows	1

In [118]:
```
plt.figure(figsize=(12,8))
sns.barplot(data=tv_india,x='listed_in',y='title')
plt.xticks(rotation=90)
plt.xlabel("Genres in India")
plt.ylabel("No of TV Shows released")
plt.title("No of TV Shows released per Genre in India")
plt.show()
```



In [119]:
```
tv_genres = tv_genre.merge(tv_genres_max,on=['country','title'],how="inner")
tv_genres
```

Out[119]:

	country	listed_in	title
0		International TV Shows	1
1		TV Dramas	1
2	Argentina	Spanish-Language TV Shows	18
3	Australia	International TV Shows	31
4	Austria	Crime TV Shows	1
...	...	...	...
105	Uruguay	International TV Shows	1
106	Uruguay	Science & Nature TV	1
107	West Germany	International TV Shows	1
108	West Germany	TV Comedies	1
109	West Germany	TV Dramas	1

110 rows × 3 columns

From above we can see the Most popular genre of TV Show for each of the country.

Best time to launch a TV Show

What is the best time to launch a new TV Show ?

In [120]:
```
TVShow.head()
```

Out[120]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration in seasons	listed_in
0	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Mysteries
3	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas

In [121]:
```
TVShow['added_month'] = TVShow['date_added'].dt.month_name()
TVShow['added_year'] = TVShow['date_added'].dt.year
TVShow
```

```
<ipython-input-121-1075ef213fe5>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  TVShow['added_month'] = TVShow['date_added'].dt.month_name()
<ipython-input-121-1075ef213fe5>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  TVShow['added_year'] = TVShow['date_added'].dt.year
```

Out[121]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration in seasons	listed_in	added_month	added_year
0	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows	September	2021.0
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas	September	2021.0
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Mysteries	September	2021.0
3	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows	September	2021.0
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas	September	2021.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
56143	s8801	TV Show	Zindagi Gulzar Hai	NaN	Hina Khawaja Bayat	Pakistan	2016-12-15	2012	TV-PG	1	Romantic TV Shows	December	2016.0
56144	s8801	TV Show	Zindagi Gulzar Hai	NaN	Hina Khawaja Bayat	Pakistan	2016-12-15	2012	TV-PG	1	TV Dramas	December	2016.0
56145	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2018	TV-Y7	2	Kids' TV	July	2019.0
56146	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2018	TV-Y7	2	Korean TV Shows	July	2019.0
56147	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2018	TV-Y7	2	TV Comedies	July	2019.0

56148 rows × 13 columns

In [122]:
```
tv_months = TVShow.groupby(['added_year','added_month'])['title'].nunique().to_frame().reset_index()
tv_month = tv_months.groupby('added_year')['title'].agg(func="max")
tv_month = tv_months.merge(tv_month,on = ['added_year','title'],how="inner")
tv_month
```

Out[122]:

	added_year	added_month	title
0	2008.0	February	1
1	2013.0	October	2
2	2014.0	November	2
3	2015.0	December	7
4	2016.0	December	44
5	2017.0	August	38
6	2017.0	December	38
7	2017.0	March	38
8	2018.0	December	61
9	2019.0	November	68
10	2020.0	December	68
11	2021.0	July	88

Over the years, more number of TV Shows were released in the month of December. As discussed earlier since its the festive season, it would be an ideal month to launch a new TV Show.

Popular director of TV Shows in different Genres

In [123]:
```
TVShow.head()
```

Out[123]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration in seasons	listed_in	added_month	added_year
0	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows	September	2021.0
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas	September	2021.0
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Mysteries	September	2021.0
3	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows	September	2021.0
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas	September	2021.0

In [124]:
```
poptv_direc = TVShow.groupby(['listed_in','director'])['title'].nunique().to_frame().reset_index()
poptv = poptv_direc.groupby('listed_in')['title'].agg(func="max").to_frame().reset_index()
poptv_direc = poptv_direc.merge(poptv,on=['listed_in','title'],how="inner")
poptv_direc
```

Out[124]:

	listed_in	director	title
0	Anime Series	Go Koga	1
1	Anime Series	Hayato Date	1
2	Anime Series	Jay Oliva	1
3	Anime Series	Kazuya Murata	1
4	Anime Series	Maite Ruiz De Austri	1
...	...	...	...
313	TV Thrillers	Weica Wang	1
314	TV Thrillers	YC Tom Lee	1
315	Teen TV Shows	Pass Patthanakumjon	1
316	Teen TV Shows	Stefan Brogren	1
317	Teen TV Shows	Takuya Igarashi	1

318 rows × 3 columns

In [125]:
```
genpop_tv = poptv_direc.groupby('listed_in')['director'].unique().to_frame().reset_index()
genpop_tv
```

Out[125]:

	listed_in	director
0	Anime Series	[Go Koga, Hayato Date, Jay Oliva, Kazuya Murat...
1	British TV Shows	[Alastair Fothergill]
2	Classic & Cult TV	[Greg Tiernan, Michael Cumming, Phil Sgriccia,...
3	Crime TV Shows	[Ahmet Katıksız, Alain Brunard, Alejandro Hart...
4	Docuseries	[Alastair Fothergill, Ken Burns]
5	International TV Shows	[Alastair Fothergill]
6	Kids' TV	[Iginio Straffi]
7	Korean TV Shows	[Jung-ah Im, Shin Won-ho]
8	Reality TV	[Adrián García Bogliano, Andy Devonshire, Bump...
9	Romantic TV Shows	[Hsu Fu-chun, Shin Won-ho]
10	Science & Nature TV	[Estela Renner, Everardo Gout, Jason Hehir, St...
11	Spanish-Language TV Shows	[Adrián García Bogliano, Alejandro Lozano, And...
12	Stand-Up Comedy & Talk Shows	[Jung-ah Im, Stan Lathan]
13	TV Action & Adventure	[Ange Basterga, DJ Chen, Danny Cannon, Gary Si...
14	TV Comedies	[Stan Lathan]
15	TV Dramas	[Abhishek Chaubey, Aco Tenriyagelli, Ahmet Kat...
16	TV Horror	[Chuang Shiang-an, Eli Roth, Hong Won-ki, Jay ...
17	TV Mysteries	[Abhishek Chaubey, Barbara Schroeder, Brad And...
18	TV Sci-Fi & Fantasy	[Glen Winter, Jesse Warn, Rob Seidenglanz, Tob...
19	TV Shows	[Ali Kalthami, Arvind Swamy, Bejoy Nambiar, Br...
20	TV Thrillers	[Chen Hung-yi, Cho Li, Eli Roth, Lin Guan-fu, ...
21	Teen TV Shows	[Pass Patthanakumjon, Stefan Brogren, Takuya I...

From the above we can see the Most popular TV Show directors for each of the Genre

Popular TV Show Actor in each country

In [126]:
```
TVShow.head()
```

Out[126]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration in seasons	listed_in	added_month	added_year
0	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows	September	2021.0
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas	September	2021.0
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Mysteries	September	2021.0
3	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows	September	2021.0
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas	September	2021.0

In [127]:
```
tv_actor = TVShow.groupby(['country','cast'])['title'].nunique().to_frame().reset_index()
tv_actor
```

Out[127]:

	country	cast	title
0		Jung Hae-in	1
1		Kim Sung-kyun	1
2		Koo Kyo-hwan	1
3		Son Suk-ku	1
4	Argentina	Abel Ayala	1
...	...	...	...
16831	West Germany	Roland Mqwebu	1
16832	West Germany	Roy Dotrice	1
16833	West Germany	Terry Jones	1
16834	West Germany	Trevor Howard	1
16835	West Germany	Tu Nokwe	1

16836 rows × 3 columns

In [128]:
```
poptvactor = tv_actor.groupby('country')['title'].agg(func="max").to_frame().reset_index()
tv_actor = tv_actor.merge(poptvactor,on=['country','title'],how='inner')
tv_actor.sort_values(by='title',ascending=False,inplace=True)
tv_actor.reset_index(drop=True,inplace=True)
tv_actor
```

Out[128]:

	country	cast	title
0	Japan	Takahiro Sakurai	22
1	United Kingdom	David Attenborough	13
2	United States	Grey Griffin	10
3	Singapore	Zhang Zhenhuan	7
4	Canada	Ashleigh Ball	7
...	...	...	...
489	Jordan	Aysha Shahaltough	1
490	Jordan	Andria Tayeh	1
491	Indonesia	Yusril Fahriza	1
492	Indonesia	Steve Blum	1
493	West Germany	Tu Nokwe	1

494 rows × 3 columns

From the above we can see the popular TV Show actor and the number of his/her movies released in netflix platform

In [129]:
```
tv_actor.loc[tv_actor['country']=='India']  #Popular TV Show actors in India
```

Out[129]:

	country	cast	title
37	India	Rajesh Kava	3
38	India	Nishka Raheja	3

For example: The Most popular TV Show actors in a country like India are: **Rajesh Kava** and **Nishka Raheja**.

Number of TV Shows under each rating

In [130]:
```
TVShow.head()
```

Out[130]:

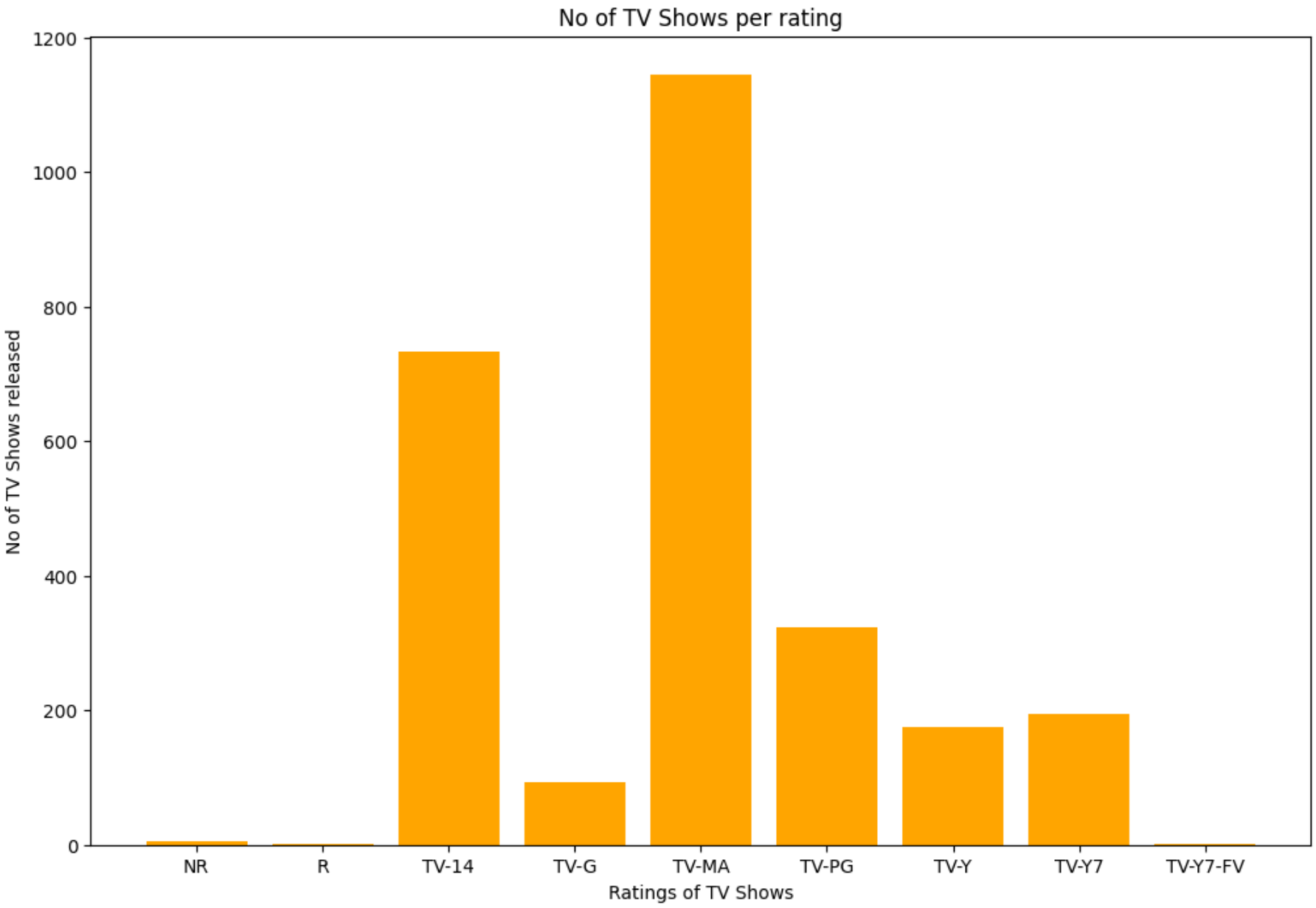
	show_id	type	title	director	cast	country	date_added	release_year	rating	duration in seasons	listed_in	added_month	added_year
0	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows	September	2021.0
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas	September	2021.0
2	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2	TV Mysteries	September	2021.0
3	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	International TV Shows	September	2021.0
4	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2	TV Dramas	September	2021.0

In [131]:
```
tv_rating = TVShow.groupby('rating')['title'].nunique().to_frame().reset_index()
tv_rating
```

Out[131]:

	rating	title
0	NR	5
1	R	2
2	TV-14	733
3	TV-G	94
4	TV-MA	1145
5	TV-PG	323
6	TV-Y	176
7	TV-Y7	195
8	TV-Y7-FV	1

In [132]:
```
plt.figure(figsize=(12,8))
plt.bar(tv_rating['rating'],tv_rating['title'],color="orange")
plt.xlabel('Ratings of TV Shows')
plt.ylabel('No of TV Shows released')
plt.title("No of TV Shows per rating")
plt.show()
```



Thus we can see there are more number of TV Shows released with the rating of TV-MA which means majority of the shows are intended only for

adults.

In [132]: