

In [1]:

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

In [2]:

```
data=pd.read_csv("C://Users//shiva//netflix.csv")  
data
```

Out[2]:

	show_id	type		title	director	cast	country	date_added	release_year
0	s1	Movie		Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2021
1	s2	TV Show		Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show		Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show		Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show		Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
...
8802	s8803	Movie		Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2006
8803	s8804	TV Show		Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2019
8804	s8805	Movie		Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009

	show_id	type	title	director	cast	country	date_added	release_year
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	200
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	March 2, 2019	201

Analyziing Basic Metrics

8807 rows × 12 columns

In [3]:

```
data.dtypes
```

Out[3]:

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

In [4]:

```
len(data)
```

Out[4]:

8807

In [5]:

```
data.head(10)
```

Out[5]:

	show_id	type	title	director	cast	country	date_added	release_yea
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2021
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	NaN	September 24, 2021	2021
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1991
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...	United Kingdom	September 24, 2021	2021

	show_id	type	title	director	cast	country	date_added	release_yea
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021

In [6]:

```
data.shape
```

Out[6]:

(8807, 12)

In [7]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   show_id               8807 non-null   object
1   type                  8807 non-null   object
2   title                 8807 non-null   object
3   director              6173 non-null   object
4   cast                  7982 non-null   object
5   country               7976 non-null   object
6   date_added            8797 non-null   object
7   release_year          8807 non-null   int64
8   rating                8803 non-null   object
9   duration              8804 non-null   object
10  listed_in             8807 non-null   object
11  description            8807 non-null   object
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
```

Changing Datatype

In [8]:

```
data["date_added"] = data["date_added"].astype('datetime64[ns]')
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   show_id         8807 non-null   object
 1   type            8807 non-null   object
 2   title           8807 non-null   object
 3   director        6173 non-null   object
 4   cast            7982 non-null   object
 5   country         7976 non-null   object
 6   date_added      8797 non-null   datetime64[ns]
 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: datetime64[ns](1), int64(1), object(10)
memory usage: 825.8+ KB
```

Checking Null Values

In [9]:

```
data.isnull().sum()/len(data)*100
```

Out[9]:

```
show_id      0.000000
type         0.000000
title        0.000000
director     29.908028
cast         9.367549
country      9.435676
date_added   0.113546
release_year 0.000000
rating       0.045418
duration     0.034064
listed_in    0.000000
description  0.000000
dtype: float64
```


In [10]:

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

Out[10]:

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

In [11]:

```
data["cast"].value_counts()
```

Out[11]:

```
David Attenborough
19
Vatsal Dubey, Julie Tejjwani, Rupa Bhimani, Jigna Bhardwaj, Rajesh Kava,
Mousam, Swapnil
14
Samuel West
10
Jeff Dunham
7
David Spade, London Hughes, Fortune Feimster
6

..
Michael Peña, Diego Luna, Tenoch Huerta, Joaquin Cosio, José María Yazp
ik, Matt Letscher, Alyssa Diaz
1
Nick Lachey, Vanessa Lachey
1
Takeru Sato, Kasumi Arimura, Haru, Kentaro Sakaguchi, Takayuki Yamada,
Kendo Kobayashi, Ken Yasuda, Arata Furuta, Suzuki Matsuo, Koichi Yamade
ra, Arata Iura, Chikako Kaku, Kotaro Yoshida      1
Toyin Abraham, Sambasa Nzeribe, Chioma Chukwuka Akpotha, Chioma Omerua
h, Chiwetalu Agu, Dele Odule, Femi Adebayo, Bayray McNwizu, Biodun Step
hen      1
Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary, Megh
na Malik, Malkeet Rauni, Anita Shabdish, Chittaranjan Tripathy
1
Name: cast, Length: 7692, dtype: int64
```

Un-Nesting

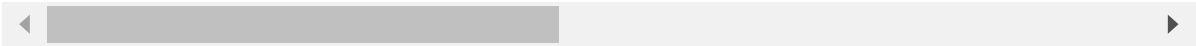
In [12]:

```
new_data=data["cast"].apply(lambda x: str(x).split(", ")).tolist()
data_new=pd.DataFrame(new_data,index=data["title"])
data_new
```

Out[12]:

	0	1	2	3	4	5	6
title							
Dick Johnson Is Dead	nan	None	None	None	None	None	None
Blood & Water	Ama Qamata	Khosi Ngema	Gail Mabalane	Thabang Molaba	Dillon Windvogel	Natasha Thahane	Arno Greeff
Ganglands	Sami Bouajila	Tracy Gotoas	Samuel Jouy	Nabiha Akkari	Sofia Lesaffre	Salim Kechiouche	Noureddin Farif
Jailbirds New Orleans	nan	None	None	None	None	None	None
Kota Factory	Mayur More	Jitendra Kumar	Ranjan Raj	Alam Khan	Ahsaas Channa	Revathi Pillai	Urvi Singh
...
Zodiac	Mark Ruffalo	Jake Gyllenhaal	Robert Downey Jr.	Anthony Edwards	Brian Cox	Elias Koteas	Don Loggins
Zombie Dumb	nan	None	None	None	None	None	None
Zombieland	Jesse Eisenberg	Woody Harrelson	Emma Stone	Abigail Breslin	Amber Heard	Bill Murray	Derek Gray
Zoom	Tim Allen	Courteney Cox	Chevy Chase	Kate Mara	Ryan Newman	Michael Cassidy	Spence Breslin
Zubaan	Vicky Kaushal	Sarah-Jane Dias	Raaghav Chanana	Manish Chaudhary	Meghna Malik	Malkeet Rauni	Anit Shabdis

8807 rows × 50 columns



In [13]:

```
df_new=data_new.stack()  
df_new=pd.DataFrame(df_new)  
df_new.reset_index(inplace=True)  
df_new
```

Out[13]:

	title	level_1	0
0	Dick Johnson Is Dead	0	nan
1	Blood & Water	0	Ama Qamata
2	Blood & Water	1	Khosi Ngema
3	Blood & Water	2	Gail Mabalane
4	Blood & Water	3	Thabang Molaba
...
64946	Zubaan	3	Manish Chaudhary
64947	Zubaan	4	Meghna Malik
64948	Zubaan	5	Malkeet Rauni
64949	Zubaan	6	Anita Shabdish
64950	Zubaan	7	Chittaranjan Tripathy

64951 rows × 3 columns

In []:

In [14]:

```
df_new.drop("level_1",axis=1,inplace=True)
```

In [15]:

```
df_new.rename(columns={0:"Actors"},inplace=True)
```

In [16]:

```
df_new
```

Out[16]:

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

64951 rows × 2 columns

In [17]:

```
data["director"].value_counts()
```

Out[17]:

Rajiv Chilaka	19
Raúl Campos, Jan Suter	18
Marcus Raboy	16
Suhas Kadav	16
Jay Karas	14
..	..
Raymie Muzquiz, Stu Livingston	1
Joe Menendez	1
Eric Bross	1
Will Eisenberg	1
Mozez Singh	1
Name: director, Length: 4528, dtype: int64	

In [18]:

```
data["director"].nunique()
```

Out[18]:

4528

In [19]:

```
director_split=data["director"].apply(lambda x: str(x).split(", ")).tolist()
director_new=pd.DataFrame(director_split,index=data["title"])
director_new=director_new.stack()
director_new=pd.DataFrame(director_new.reset_index())
director_new
```

Out[19]:

	title	level_1	0
0	Dick Johnson Is Dead	0	Kirsten Johnson
1	Blood & Water	0	nan
2	Ganglands	0	Julien Leclercq
3	Jailbirds New Orleans	0	nan
4	Kota Factory	0	nan
...
9607	Zodiac	0	David Fincher
9608	Zombie Dumb	0	nan
9609	Zombieland	0	Ruben Fleischer
9610	Zoom	0	Peter Hewitt
9611	Zubaan	0	Mozes Singh

9612 rows × 3 columns

In [20]:

```
director_new.rename(columns={0:"Directors"},inplace=True)
director_new.drop("level_1",axis=1,inplace=True)
```

In [21]:

```
director_new
```

Out[21]:

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

9612 rows × 2 columns

In [22]:

```
data["country"].nunique()
```

Out[22]:

748

In [23]:

```
data["country"].value_counts()
```

Out[23]:

United States	2818
India	972
United Kingdom	419
Japan	245
South Korea	199
...	
Romania, Bulgaria, Hungary	1
Uruguay, Guatemala	1
France, Senegal, Belgium	1
Mexico, United States, Spain, Colombia	1
United Arab Emirates, Jordan	1
Name: country, Length: 748, dtype: int64	

In [24]:

```
country_split=data["country"].apply(lambda x:str(x).split(", ")).tolist()
```

In [25]:

```
country_new=pd.DataFrame(country_split,index=data["title"])
```

In [26]:

```
country_new=country_new.stack()
country_new=pd.DataFrame(country_new.reset_index())
country_new
```

Out[26]:

	title	level_1	0
0	Dick Johnson Is Dead	0	United States
1	Blood & Water	0	South Africa
2	Ganglands	0	nan
3	Jailbirds New Orleans	0	nan
4	Kota Factory	0	India
...
10840	Zodiac	0	United States
10841	Zombie Dumb	0	nan
10842	Zombieland	0	United States
10843	Zoom	0	United States
10844	Zubaan	0	India

10845 rows × 3 columns

In [27]:

```
country_new.drop("level_1",axis=1,inplace=True)
```

In [28]:

```
country_new.rename(columns={0:"Country"},inplace=True)
```

In [29]:

country_new

Out[29]:

	title	Country
0	Dick Johnson Is Dead	United States
1	Blood & Water	South Africa
2	Ganglands	nan
3	Jailbirds New Orleans	nan
4	Kota Factory	India
...
10840	Zodiac	United States
10841	Zombie Dumb	nan
10842	Zombieland	United States
10843	Zoom	United States
10844	Zubaan	India

10845 rows × 2 columns

In [30]:

```
listed_split=data["listed_in"].apply(lambda x:str(x).split(", ")).tolist()
listed_split
```

Out[30]:

```
[['Documentaries'],
 ['International TV Shows', 'TV Dramas', 'TV Mysteries'],
 ['Crime TV Shows', 'International TV Shows', 'TV Action & Adventure'],
 ['Docuseries', 'Reality TV'],
 ['International TV Shows', 'Romantic TV Shows', 'TV Comedies'],
 ['TV Dramas', 'TV Horror', 'TV Mysteries'],
 ['Children & Family Movies'],
 ['Dramas', 'Independent Movies', 'International Movies'],
 ['British TV Shows', 'Reality TV'],
 ['Comedies', 'Dramas'],
 ['Crime TV Shows', 'Docuseries', 'International TV Shows'],
 ['Crime TV Shows', 'International TV Shows', 'TV Action & Adventure'],
 ['Dramas', 'International Movies'],
 ['Children & Family Movies', 'Comedies'],
 ['British TV Shows', 'Crime TV Shows', 'Docuseries'],
 ['TV Comedies', 'TV Dramas']]
```


In [31]:

```
Category_new=pd.DataFrame(listed_split,index=data["title"])
Category_new
```

Out[31]:

	0	1	2
title			
Dick Johnson Is Dead	Documentaries	None	None
Blood & Water	International TV Shows	TV Dramas	TV Mysteries
Ganglands	Crime TV Shows	International TV Shows	TV Action & Adventure
Jailbirds New Orleans	Docuseries	Reality TV	None
Kota Factory	International TV Shows	Romantic TV Shows	TV Comedies
...
Zodiac	Cult Movies	Dramas	Thrillers
Zombie Dumb	Kids' TV	Korean TV Shows	TV Comedies
Zombieland	Comedies	Horror Movies	None
Zoom	Children & Family Movies	Comedies	None
Zubaan	Dramas	International Movies	Music & Musicals

8807 rows × 3 columns

In [32]:

```
Category_new=Category_new.stack()  
Category_new=pd.DataFrame(Category_new.reset_index())  
Category_new
```

Out[32]:

	title	level_1	0
0	Dick Johnson Is Dead	0	Documentaries
1	Blood & Water	0	International TV Shows
2	Blood & Water	1	TV Dramas
3	Blood & Water	2	TV Mysteries
4	Ganglands	0	Crime TV Shows
...
19318	Zoom	0	Children & Family Movies
19319	Zoom	1	Comedies
19320	Zubaan	0	Dramas
19321	Zubaan	1	International Movies
19322	Zubaan	2	Music & Musicals

19323 rows × 3 columns

In [33]:

```
Category_new.drop("level_1",axis=1,inplace=True)
Category_new.rename(columns={0:"Category"},inplace=True)
Category_new
```

Out[33]:

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

19323 rows × 2 columns

Merging DataFrames

In [34]:

```
df_latest=df_new.merge(director_new,on=["title"],how="inner")
df_latest1=df_latest.merge(Category_new,on=["title"],how="inner")
df_latest2=df_latest1.merge(country_new,on=["title"],how="inner")
df_latest2
```

Out[34]:

	title	Actors	Directors	Category	Country
0	Dick Johnson Is Dead	nan	Kirsten Johnson	Documentaries	United States
1	Blood & Water	Ama Qamata	nan	International TV Shows	South Africa
2	Blood & Water	Ama Qamata	nan	TV Dramas	South Africa
3	Blood & Water	Ama Qamata	nan	TV Mysteries	South Africa
4	Blood & Water	Khosi Ngema	nan	International TV Shows	South Africa
...
201986	Zubaan	Anita Shabdish	Mozez Singh	International Movies	India
201987	Zubaan	Anita Shabdish	Mozez Singh	Music & Musicals	India
201988	Zubaan	Chittaranjan Tripathy	Mozez Singh	Dramas	India
201989	Zubaan	Chittaranjan Tripathy	Mozez Singh	International Movies	India
201990	Zubaan	Chittaranjan Tripathy	Mozez Singh	Music & Musicals	India

201991 rows × 5 columns

In [35]:

```
df_latest2.isnull().sum()
```

Out[35]:

```
title      0
Actors     0
Directors  0
Category   0
Country    0
dtype: int64
```

In [36]:

```
df_latest2["Actors"].replace(['nan'],["Actor Unknown"],inplace=True)
df_latest2["Directors"].replace(['nan'],['Director Unknown'],inplace=True)
df_latest2["Country"].replace(["nan"],[np.nan],inplace=True)
df_latest2.head(10)
```

Out[36]:

	title	Actors	Directors	Category	Country
0	Dick Johnson Is Dead	Actor Unknown	Kirsten Johnson	Documentaries	United States
1	Blood & Water	Ama Qamata	Director Unknown	International TV Shows	South Africa
2	Blood & Water	Ama Qamata	Director Unknown	TV Dramas	South Africa
3	Blood & Water	Ama Qamata	Director Unknown	TV Mysteries	South Africa
4	Blood & Water	Khosi Ngema	Director Unknown	International TV Shows	South Africa
5	Blood & Water	Khosi Ngema	Director Unknown	TV Dramas	South Africa
6	Blood & Water	Khosi Ngema	Director Unknown	TV Mysteries	South Africa
7	Blood & Water	Gail Mabalane	Director Unknown	International TV Shows	South Africa
8	Blood & Water	Gail Mabalane	Director Unknown	TV Dramas	South Africa
9	Blood & Water	Gail Mabalane	Director Unknown	TV Mysteries	South Africa

In [37]:

```
df_latest2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 201991 entries, 0 to 201990
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   title       201991 non-null object
1   Actors      201991 non-null object
2   Directors   201991 non-null object
3   Category    201991 non-null object
4   Country     190094 non-null object
dtypes: object(5)
memory usage: 9.2+ MB
```

In [38]:

```
dflatest_3=df_latest2.merge(data[['show_id', 'type', 'title', 'date_added',  
    'release_year', 'rating', 'duration']],on=["title"],how="left")  
dflatest_3
```

Out[38]:

	title	Actors	Directors	Category	Country	show_id	type	date_ad
0	Dick Johnson Is Dead	Actor Unknown	Kirsten Johnson	Documentaries	United States	s1	Movie	2021-0
1	Blood & Water	Ama Qamata	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-0
2	Blood & Water	Ama Qamata	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-0
3	Blood & Water	Ama Qamata	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-0
4	Blood & Water	Khosi Ngema	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-0
...
201986	Zubaan	Anita Shabdish	Mozez Singh	International Movies	India	s8807	Movie	2019-0
201987	Zubaan	Anita Shabdish	Mozez Singh	Music & Musicals	India	s8807	Movie	2019-0
201988	Zubaan	Chittaranjan Tripathy	Mozez Singh	Dramas	India	s8807	Movie	2019-0
201989	Zubaan	Chittaranjan Tripathy	Mozez Singh	International Movies	India	s8807	Movie	2019-0
201990	Zubaan	Chittaranjan Tripathy	Mozez Singh	Music & Musicals	India	s8807	Movie	2019-0

201991 rows × 11 columns



In [39]:

```
dflatest_3.isnull().sum()
```

Out[39]:

```
title           0
Actors          0
Directors       0
Category        0
Country        11897
show_id        0
type           0
date_added     158
release_year    0
rating         67
duration        3
dtype: int64
```

In [40]:

```
dflatest_3["duration"].value_counts()
```

Out[40]:

```
1 Season      35035
2 Seasons     9559
3 Seasons     5084
94 min        4343
106 min       4040
...
3 min         4
5 min         3
11 min        2
8 min         2
9 min         2
Name: duration, Length: 220, dtype: int64
```

In [41]:

```
dflatest_3["rating"].value_counts()
```

Out[41]:

```
TV-MA      73867
TV-14      43931
R           25860
PG-13      16246
TV-PG      14926
PG          10919
TV-Y7       6304
TV-Y        3665
TV-G        2779
NR           1573
G            1530
NC-17        149
TV-Y7-FV      86
UR            86
74 min         1
84 min         1
66 min         1
Name: rating, dtype: int64
```

In [42]:

```
dflatest_3[dflatest_3["duration"].isnull()]
```

Out[42]:

	title	Actors	Directors	Category	Country	show_id	type	date_added	release_date
126537	Louis C.K. 2017	Louis C.K.	Louis C.K.	Movies	United States	s5542	Movie	2017-04-04	
131603	Louis C.K.: Hilarious	Louis C.K.	Louis C.K.	Movies	United States	s5795	Movie	2016-09-16	
131737	Louis C.K.: Live at the Comedy Store	Louis C.K.	Louis C.K.	Movies	United States	s5814	Movie	2016-08-15	

In [43]:

```
dflatest_3.loc[dflatest_3["duration"].isnull(), "duration"] = dflatest_3.loc[dflatest_3["duration"].isnull(), "duration"].fillna(0)
```


In [44]:

```
dflatest_3.isnull().sum()
```

Out[44]:

```
title          0
Actors         0
Directors      0
Category       0
Country       11897
show_id        0
type           0
date_added     158
release_year   0
rating         67
duration       0
dtype: int64
```

In [45]:

```
dflatest_3.loc[dflatest_3["rating"].str.contains("min",na=False),"rating"]="NR"
dflatest_3["rating"].fillna("NR",inplace=True)
dflatest_3.isnull().sum()
```

Out[45]:

```
title          0
Actors         0
Directors      0
Category       0
Country       11897
show_id        0
type           0
date_added     158
release_year   0
rating         0
duration       0
dtype: int64
```

Imputing Countries -Missing Values

In [46]:

```
for i in dflatest_3[dflatest_3["Country"].isnull()]["Directors"].unique():
    if i in dflatest_3[~dflatest_3["Country"].isnull()]["Directors"].unique():
        imp=dflatest_3[dflatest_3["Directors"]==i]["Country"].mode().values[0]
        dflatest_3.loc[dflatest_3["Directors"]==i,"Country"]=dflatest_3.loc[dflatest_3["Directors"]==i,"Country"].fillna(imp)
```

In [47]:

```
for i in dflatest_3[dflatest_3["Country"].isnull()][ "Actors"].unique():
    if i in dflatest_3[~dflatest_3["Country"].isnull()][ "Actors"].unique():
        imp=dflatest_3[dflatest_3["Actors"]==i][ "Country"].mode().values[0]
        dflatest_3.loc[dflatest_3["Actors"]==i, "Country"]=dflatest_3.loc[dflatest_3["Actors"]==i, "Country"].fillna(imp)
```

Imputing date_added -Missing Values

In [48]:

```
for i in dflatest_3[dflatest_3["date_added"].isnull()][ "release_year"].unique():
    imp=dflatest_3[dflatest_3["release_year"]==i][ "date_added"].mode().values[0]
    dflatest_3.loc[dflatest_3["release_year"]==i, "date_added"]=dflatest_3.loc[dflatest_3["release_year"]==i, "date_added"].fillna(imp)
```

Checking For null Values

In [49]:

```
dflatest_3.isnull().sum()
```

Out[49]:

title	0
Actors	0
Directors	0
Category	0
Country	2069
show_id	0
type	0
date_added	0
release_year	0
rating	0
duration	0
dtype:	int64

In [50]:

```
dflatest_3.fillna("Country Not Found",inplace=True)
```

In [51]:

```
dflatest_3.isnull().sum()
```

Out[51]:

```
title          0
Actors         0
Directors      0
Category       0
Country        0
show_id        0
type           0
date_added     0
release_year   0
rating         0
duration       0
dtype: int64
```

Removing Duplicates

In [52]:

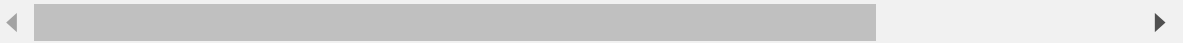
```
dflatest_3.drop_duplicates(inplace=True)
```

In [53]:

```
dflatest_3.head(10)
```

Out[53]:

	title	Actors	Directors	Category	Country	show_id	type	date_added	re
0	Dick Johnson Is Dead	Actor Unknown	Kirsten Johnson	Documentaries	United States	s1	Movie	2021-09-25	
1	Blood & Water	Ama Qamata	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	
2	Blood & Water	Ama Qamata	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-09-24	
3	Blood & Water	Ama Qamata	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-09-24	
4	Blood & Water	Khosi Ngema	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	
5	Blood & Water	Khosi Ngema	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-09-24	
6	Blood & Water	Khosi Ngema	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-09-24	
7	Blood & Water	Gail Mabalane	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	
8	Blood & Water	Gail Mabalane	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-09-24	
9	Blood & Water	Gail Mabalane	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-09-24	



In [54]:

```
dflatest_3["duration"]=dflatest_3["duration"].str.replace(" min", "")
dflatest_3.head(10)
```

Out[54]:

	title	Actors	Directors	Category	Country	show_id	type	date_added	re
0	Dick Johnson Is Dead	Actor Unknown	Kirsten Johnson	Documentaries	United States	s1	Movie	2021-09-25	
1	Blood & Water	Ama Qamata	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	
2	Blood & Water	Ama Qamata	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-09-24	
3	Blood & Water	Ama Qamata	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-09-24	
4	Blood & Water	Khosi Ngema	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	
5	Blood & Water	Khosi Ngema	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-09-24	
6	Blood & Water	Khosi Ngema	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-09-24	
7	Blood & Water	Gail Mabalane	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	
8	Blood & Water	Gail Mabalane	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-09-24	
9	Blood & Water	Gail Mabalane	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-09-24	



In [55]:

```
dflatest_3[dflatest_3["type"]=="TV Show"]["duration"].value_counts()
```

Out[55]:

```
1 Season      35035
2 Seasons     9559
3 Seasons     5084
4 Seasons     2134
5 Seasons     1698
7 Seasons      843
6 Seasons      633
8 Seasons      286
9 Seasons      257
10 Seasons     220
13 Seasons     132
12 Seasons     111
15 Seasons      96
17 Seasons      30
11 Seasons      30
Name: duration, dtype: int64
```

In [56]:

```
dflatest_3["type"].value_counts()
```

Out[56]:

```
Movie      145788
TV Show    56148
Name: type, dtype: int64
```

In [57]:

```
dflatest_3['duration_copy']=dflatest_3['duration'].copy()
df_final=dflatest_3.copy()
```

In [58]:

```
df_final.loc[df_final['duration_copy'].str.contains('Season'),'duration_copy']=0
```

In [59]:

```
df_final['duration_copy']=df_final['duration_copy'].astype('int')
df_final.head()
```

Out[59]:

	title	Actors	Directors	Category	Country	show_id	type	date_added	re
0	Dick Johnson Is Dead	Actor Unknown	Kirsten Johnson	Documentaries	United States	s1	Movie	2021-09-25	
1	Blood & Water	Ama Qamata	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	
2	Blood & Water	Ama Qamata	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-09-24	
3	Blood & Water	Ama Qamata	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-09-24	
4	Blood & Water	Khosi Ngema	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	

In [60]:

```
bins1 = [-1,1,50,80,100,120,150,200,315]
labels1 = ['<1','1-50','50-80','80-100','100-120','120-150','150-200','200-315']
df_final['duration_copy'] = pd.cut(df_final['duration_copy'],bins=bins1,labels=labels1)
df_final.head()
```

Out[60]:

	title	Actors	Directors	Category	Country	show_id	type	date_added	re
0	Dick Johnson Is Dead	Actor Unknown	Kirsten Johnson	Documentaries	United States	s1	Movie	2021-09-25	
1	Blood & Water	Ama Qamata	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	
2	Blood & Water	Ama Qamata	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-09-24	
3	Blood & Water	Ama Qamata	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-09-24	
4	Blood & Water	Khosi Ngema	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	

In [61]:

```
df_final.loc[~df_final["duration"].str.contains("Season"),"duration"]
```

Out[61]:

```
0          90
159        91
160        91
161        91
162        91
...
201986    111
201987    111
201988    111
201989    111
201990    111
Name: duration, Length: 145788, dtype: object
```

In [62]:

```
df_final.loc[~df_final["duration"].str.contains("Season"),"duration"]=df_final["durat
```

In [63]:

```
df_final.head()
```

Out[63]:

	title	Actors	Directors	Category	Country	show_id	type	date_added	re
0	Dick Johnson Is Dead	Actor Unknown	Kirsten Johnson	Documentaries	United States	s1	Movie	2021-09-25	
1	Blood & Water	Ama Qamata	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	
2	Blood & Water	Ama Qamata	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-09-24	
3	Blood & Water	Ama Qamata	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-09-24	
4	Blood & Water	Khosi Ngema	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	

In [64]:

```
df_final.drop("duration_copy",axis=1,inplace=True)
```


In [65]:

```
df_final.head()
```

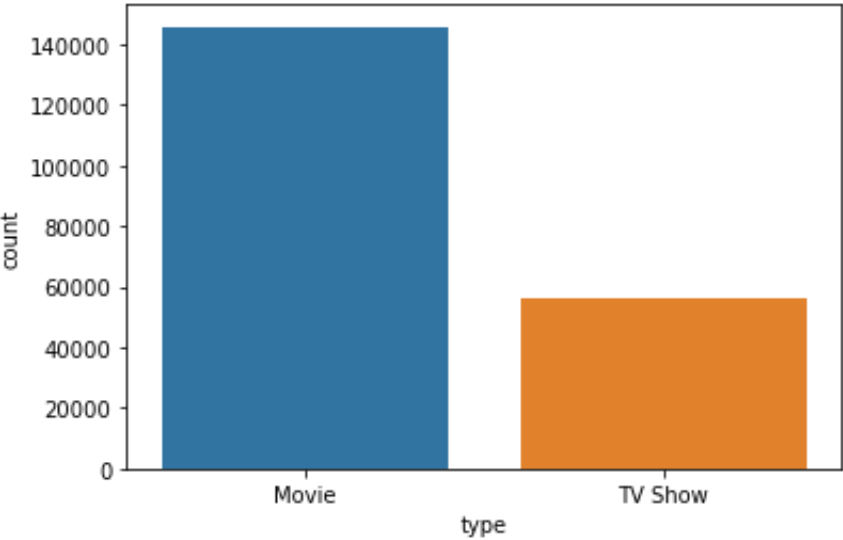
Out[65]:

	title	Actors	Directors	Category	Country	show_id	type	date_added	re
0	Dick Johnson Is Dead	Actor Unknown	Kirsten Johnson	Documentaries	United States	s1	Movie	2021-09-25	
1	Blood & Water	Ama Qamata	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	
2	Blood & Water	Ama Qamata	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-09-24	
3	Blood & Water	Ama Qamata	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-09-24	
4	Blood & Water	Khosi Ngema	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	

UNIVARIATE ANALYSIS

In [66]:

```
sns.countplot(x=df_final["type"],order=df_final["type"].value_counts().index)
plt.show()
```



In [67]:

```
df_final["release_year"].value_counts().sort_values(ascending=False)[:10]
```

Out[67]:

2018	24413
2019	21883
2017	20516
2020	19679
2016	18465
2015	14128
2021	11894
2014	9096
2013	7745
2012	6354

Name: release_year, dtype: int64

In [68]:

```
df_final["rating"].value_counts()
```

Out[68]:

TV-MA	73819
TV-14	43925
R	25859
PG-13	16246
TV-PG	14926
PG	10919
TV-Y7	6304
TV-Y	3665
TV-G	2779
NR	1643
G	1530
NC-17	149
TV-Y7-FV	86
UR	86

Name: rating, dtype: int64

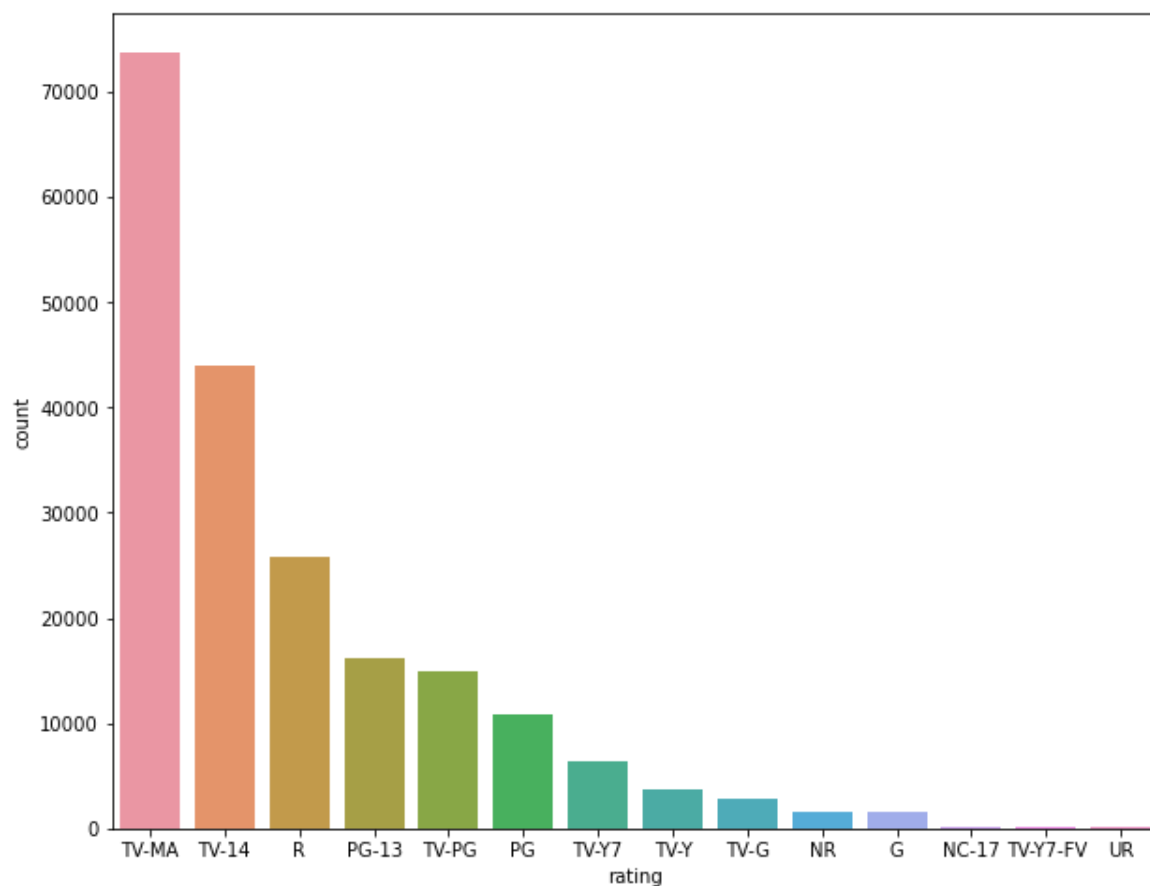
TOP 10 RATINGS

In [69]:

```
plt.figure(figsize=(10,8))
sns.countplot(x=df_final["rating"], order= df_final["rating"].value_counts().index)
```

Out[69]:

<AxesSubplot:xlabel='rating', ylabel='count'>



1. Ratings of **TV-MA** & **TV-14** & **R** are at the top

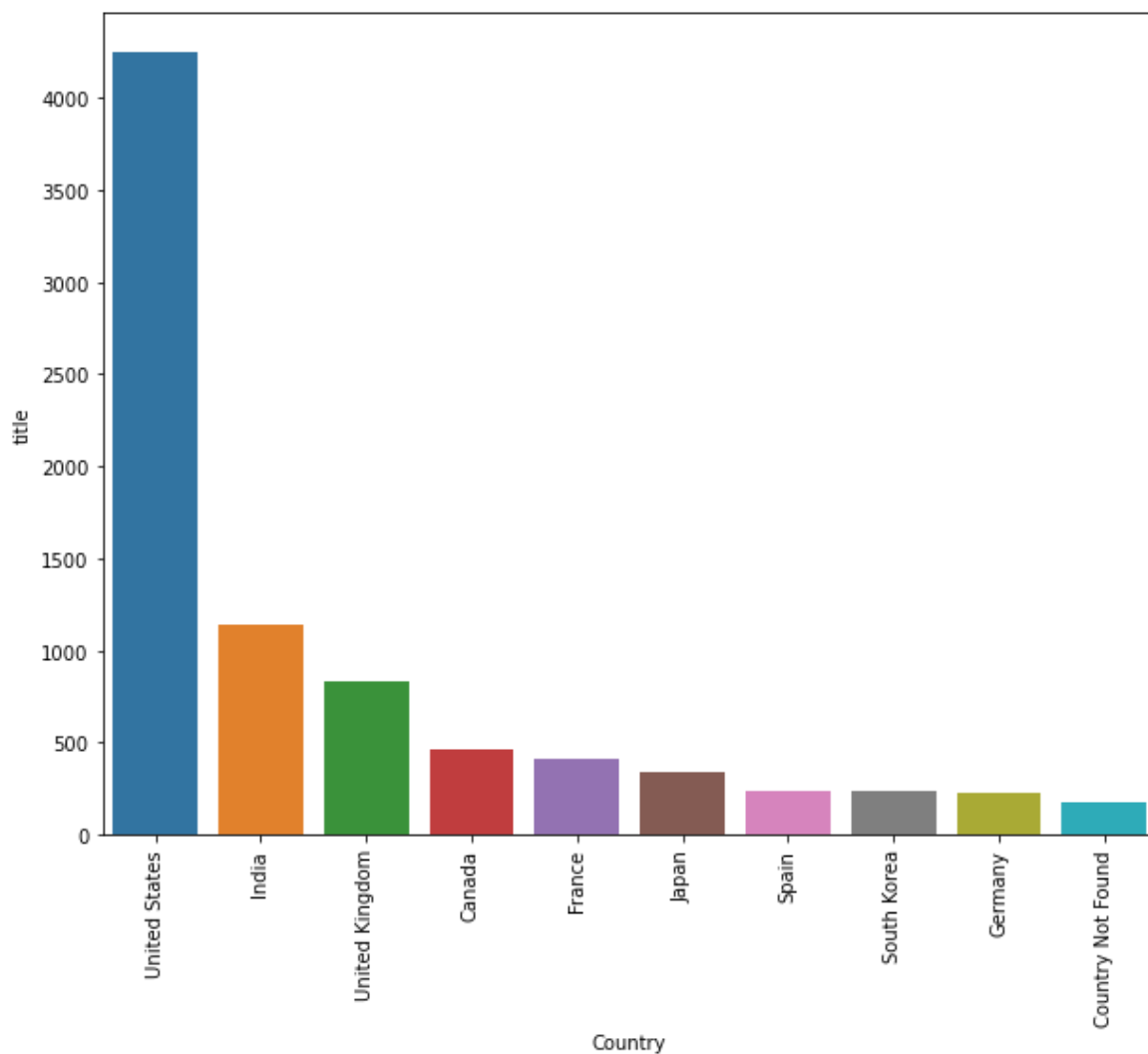
TOP 10 COUNTRIES

In [70]:

```
a=df_final.groupby(["Country"]).agg({"title": "nunique"})
a.reset_index(inplace=True)
a=a.sort_values(by=["title"],ascending=False)[:10]
```

In [71]:

```
plt.figure(figsize=(10,8))
sns.barplot(x=a["Country"],y=a["title"])
plt.xticks(rotation=90)
plt.show()
```



TOP 10 CATEGORIES

In [72]:

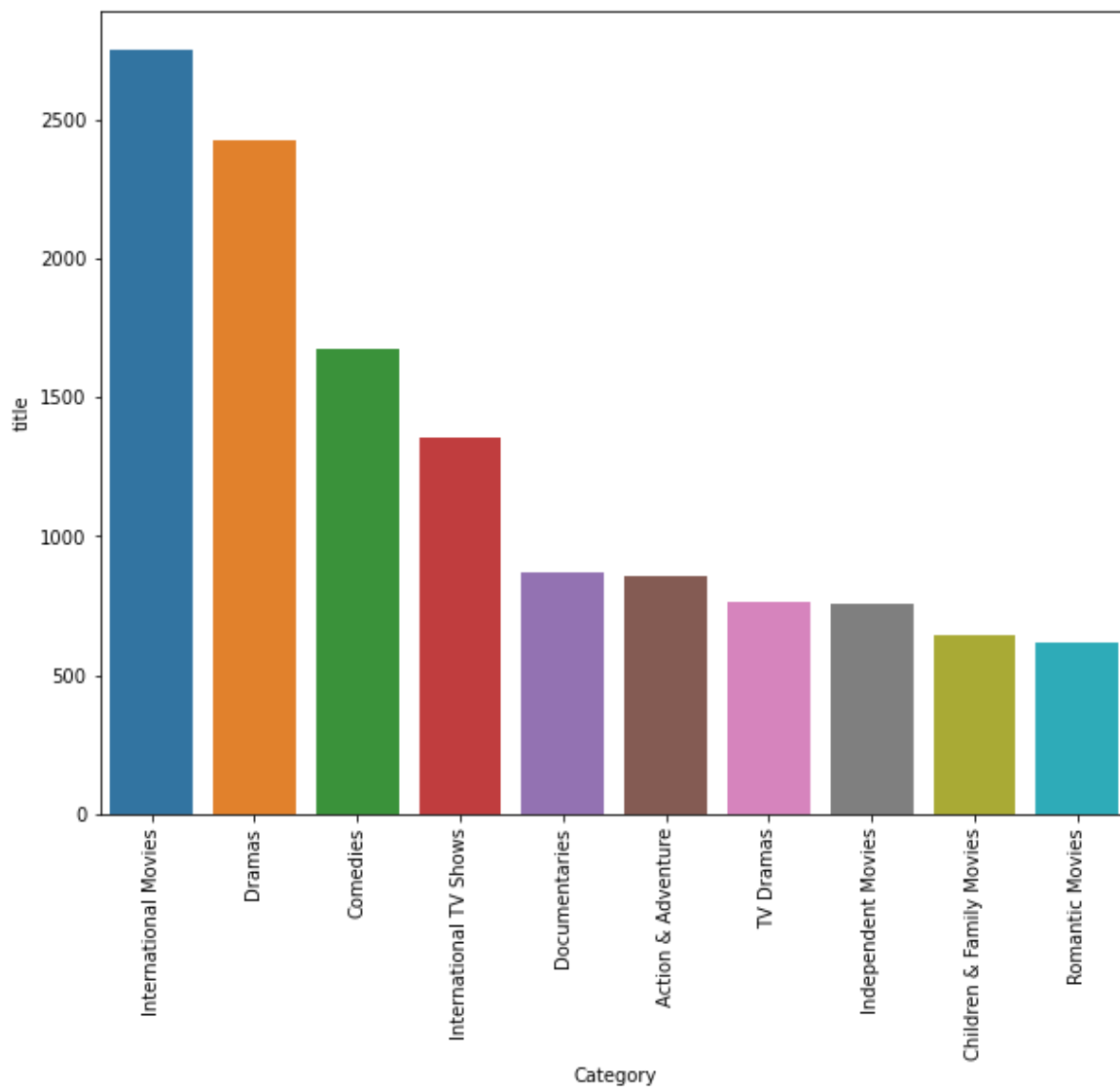
```
b=df_final.groupby(["Category"]).agg({"title":"nunique"}).sort_values(by=["title"],as
b.reset_index(inplace=True)
b
```

Out[72]:

	Category	title
0	International Movies	2752
1	Dramas	2427
2	Comedies	1674
3	International TV Shows	1351
4	Documentaries	869
5	Action & Adventure	859
6	TV Dramas	763
7	Independent Movies	756
8	Children & Family Movies	641
9	Romantic Movies	616

In [73]:

```
plt.figure(figsize=(10,8))
sns.barplot(x=b["Category"],y=b["title"])
plt.xticks(rotation=90)
plt.show()
```



1. International Movies > Dramas > Comedies are the most preferred Categories in the platform

BOTTOM 10 CATEGORIES

In [74]:

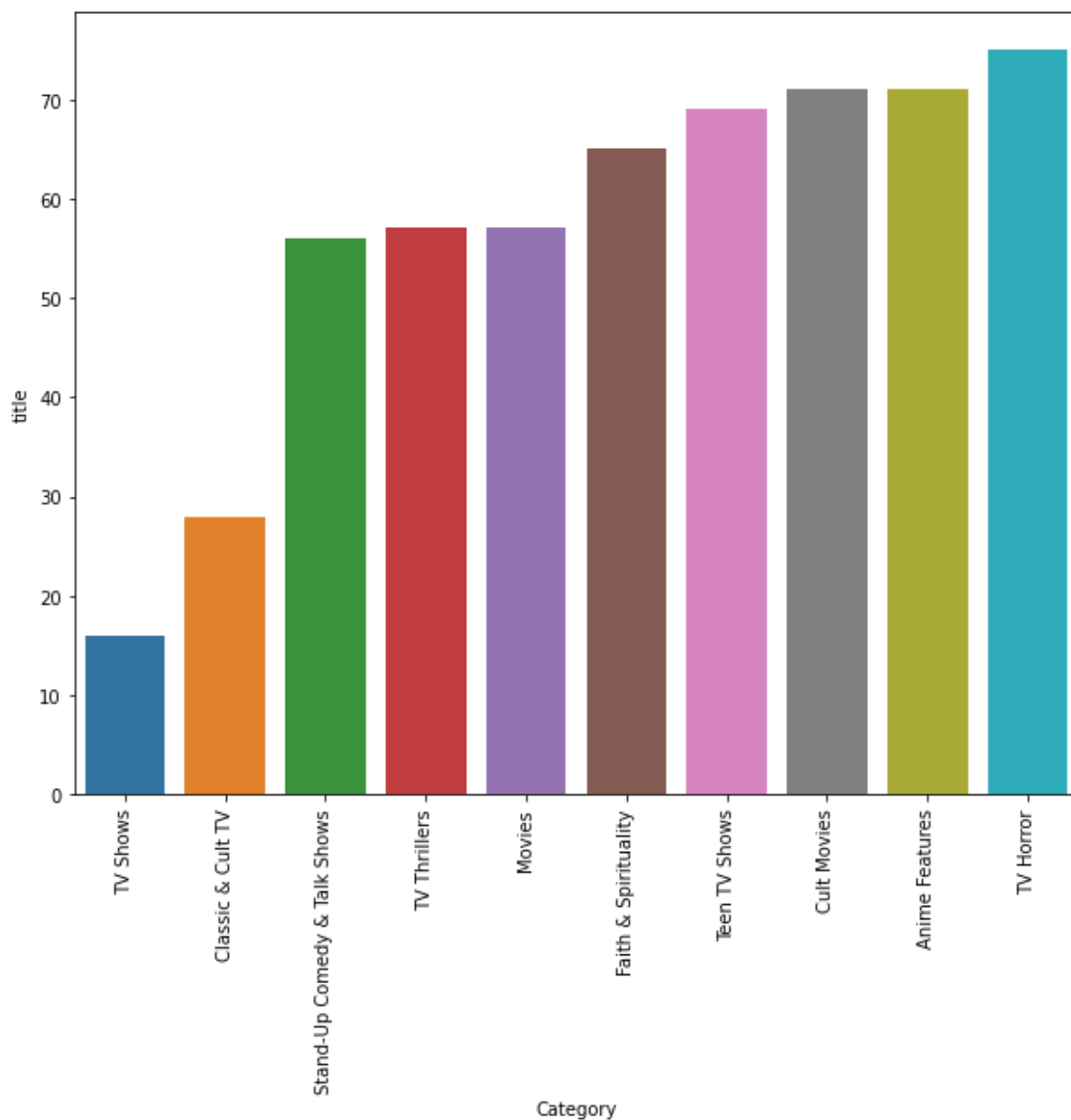
```
c=df_final.groupby(["Category"]).agg({"title":"nunique"}).sort_values(by=["title"]):
c.reset_index(inplace=True)
c
```

Out[74]:

	Category	title
0	TV Shows	16
1	Classic & Cult TV	28
2	Stand-Up Comedy & Talk Shows	56
3	TV Thrillers	57
4	Movies	57
5	Faith & Spirituality	65
6	Teen TV Shows	69
7	Cult Movies	71
8	Anime Features	71
9	TV Horror	75

In [75]:

```
plt.figure(figsize=(10,8))
sns.barplot(x=c["Category"],y=c["title"])
plt.xticks(rotation=90)
plt.show()
```



1. Classic & Cult tv shows , standups thriller are the least watched titles and categories !!

In [76]:

```
df_final["Month"]=df_final["date_added"].dt.strftime('%b')
df_final["Month"]
```

Out[76]:

```
0      Sep
1      Sep
2      Sep
3      Sep
4      Sep
...
201986  Mar
201987  Mar
201988  Mar
201989  Mar
201990  Mar
Name: Month, Length: 201936, dtype: object
```

In [77]:

```
d=df_final.groupby(["Month"]).agg({"title":"nunique"}).reset_index().sort_values(by="d")
```

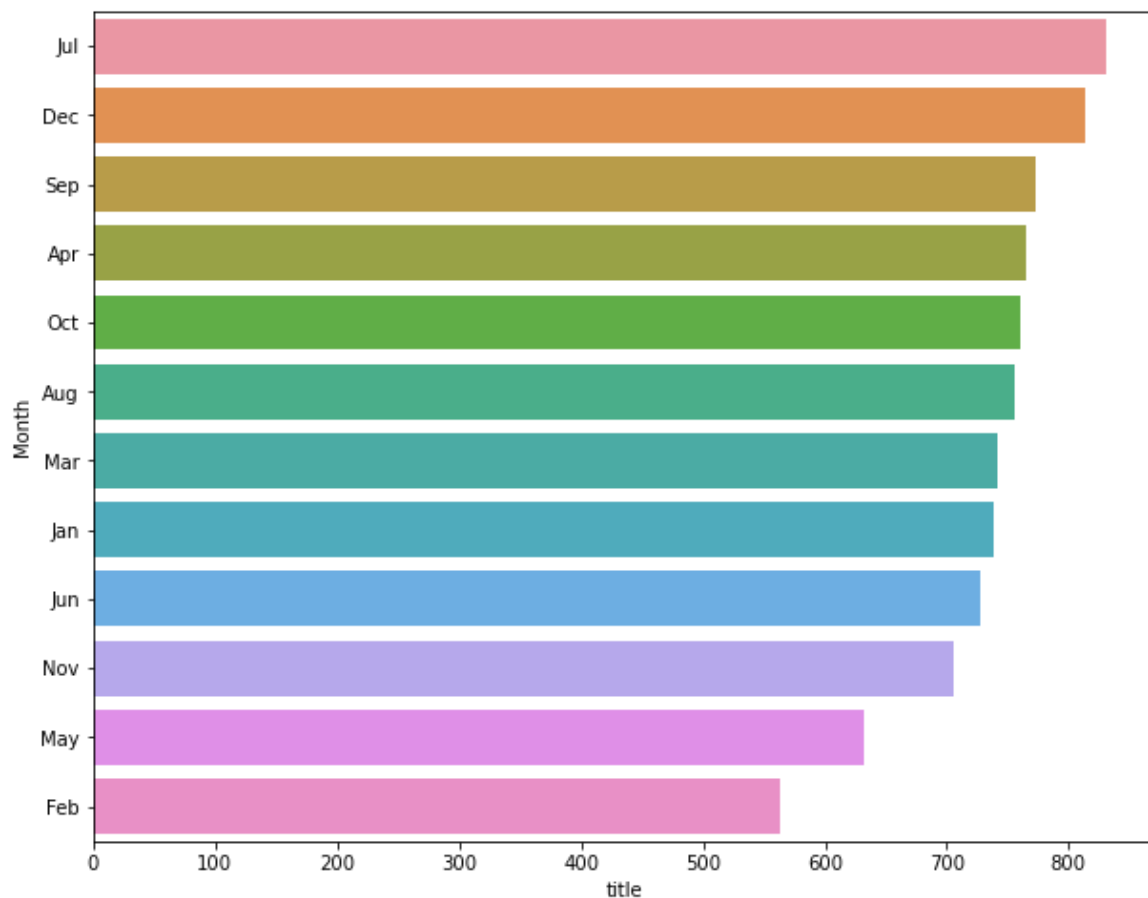
Out[77]:

	Month	title
5	Jul	831
2	Dec	813
11	Sep	773
0	Apr	765
10	Oct	760
1	Aug	756
7	Mar	742
4	Jan	738
6	Jun	728
9	Nov	706
8	May	632
3	Feb	563

MONTHLY FREQUENCY OF MOVIES ADDED

In [78]:

```
plt.figure(figsize=(10,8))
sns.barplot(y=d["Month"],x=d["title"],orient="h")
plt.show()
```



1. Mostly **July** , followed by **December** followed by **Septemeber** are the months when most titles are added !! m

MOST POPULAR SHOW : - SNOWDEN

In [79]:

```
df_final["show_id"].value_counts()
```

Out[79]:

```
s7165    700
s6985    504
s7516    468
s2554    416
s5306    378
...
s8174     1
s8176     1
s937      1
s3387     1
s1        1
Name: show_id, Length: 8807, dtype: int64
```

In [80]:

```
df_final[df_final["show_id"]=="s1765"]["title"].unique()
```

Out[80]:

```
array(['Snowden'], dtype=object)
```

In [81]:

```
df_final["Year"]=df_final["date_added"].dt.strftime("%Y")
df_final["Year"]
```

Out[81]:

```
0      2021
1      2021
2      2021
3      2021
4      2021
...
201986  2019
201987  2019
201988  2019
201989  2019
201990  2019
Name: Year, Length: 201936, dtype: object
```

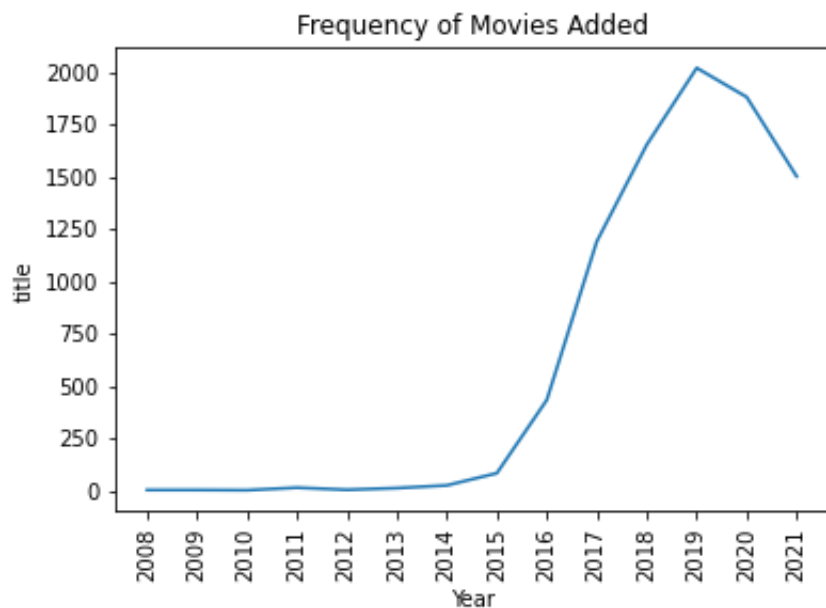
In [82]:

```
e=df_final.groupby(["Year"]).agg({"title":"nunique"})
e.reset_index(inplace=True)
```

YEARLY FREQUENCY OF MOVIES ADDED

In [83]:

```
sns.lineplot(x=e["Year"],y=e["title"])
plt.xticks(rotation=90)
plt.title("Frequency of Movies Added")
plt.show()
```



1. A **rising trend** is visible over the years in terms of **Movies** being added to platform

In [84]:

```
df_final.head()
```

Out[84]:

	title	Actors	Directors	Category	Country	show_id	type	date_added	re
0	Dick Johnson Is Dead	Actor Unknown	Kirsten Johnson	Documentaries	United States	s1	Movie	2021-09-25	
1	Blood & Water	Ama Qamata	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	
2	Blood & Water	Ama Qamata	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-09-24	
3	Blood & Water	Ama Qamata	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-09-24	
4	Blood & Water	Khosi Ngema	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-09-24	

In [85]:

```
df_final.shape
```

Out[85]:

(201936, 13)

In [86]:

```
t=df_final.groupby("duration").agg({"title":"nunique"})
t.reset_index(inplace=True)
t
```

Out[86]:

	duration	title
0	1 Season	1793
1	1-50	287
2	10 Seasons	7
3	100-120	1672
4	11 Seasons	2
5	12 Seasons	2
6	120-150	897
7	13 Seasons	3
8	15 Seasons	2
9	150-200	226
10	17 Seasons	1
11	2 Seasons	425
12	200-315	19
13	3 Seasons	199
14	4 Seasons	95
15	5 Seasons	65
16	50-80	808
17	6 Seasons	33
18	7 Seasons	23
19	8 Seasons	17
20	80-100	2222
21	9 Seasons	9

In [87]:

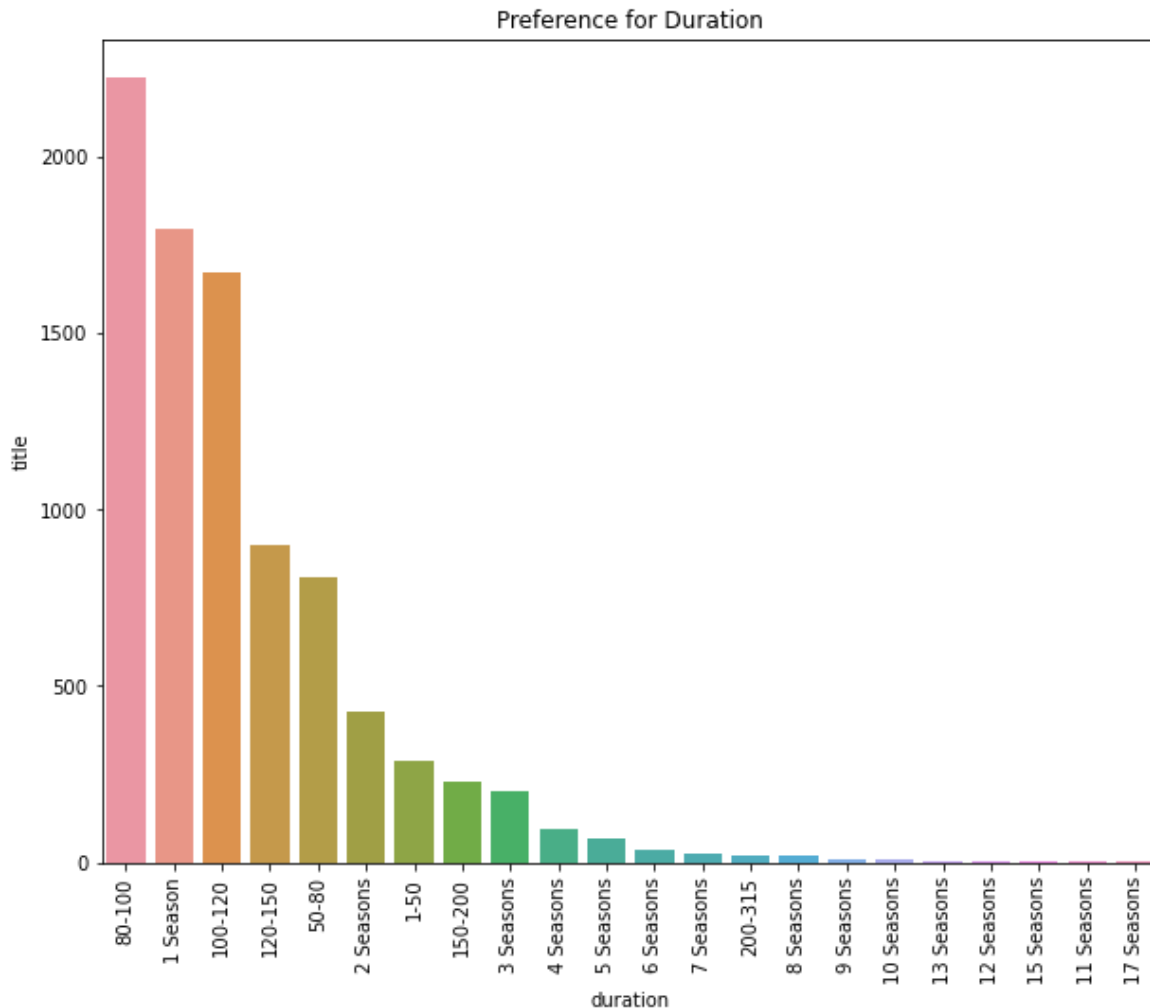
```
t.sort_values(by=["title"],ascending=False,inplace=True)
t[:10]
```

Out[87]:

	duration	title
20	80-100	2222
0	1 Season	1793
3	100-120	1672
6	120-150	897
16	50-80	808
11	2 Seasons	425
1	1-50	287
9	150-200	226
13	3 Seasons	199
14	4 Seasons	95

In [88]:

```
plt.figure(figsize=(10,8))
sns.barplot(x=t["duration"],y=t["title"])
plt.xticks(rotation=90)
plt.title("Preference for Duration")
plt.show()
```



1. Preference is usually skewed towards watching movies with 80-100 minutes duration.
2. Mostly people watch 1 season , we can derive that new series are much watched .
3. There is a general trend of people watching **Movies** in the range of 80-150 Minutes

BI VARIATE & MULTIVARIATE ANALYSIS

In [89]:

```
a=pd.crosstab(index=df_final["Category"],columns=df_final["type"],values=df_final["ti  
a
```

Out[89]:

	type	Movie	TV Show
Category			
International Movies		2752.0	NaN
Dramas		2427.0	NaN
Comedies		1674.0	NaN

In [90]:

```
b=pd.crosstab(index=df_final["Category"],columns=df_final["type"],values=df_final["ti  
b
```

Out[90]:

	type	Movie	TV Show
Category			
International TV Shows		NaN	1351.0
TV Dramas		NaN	763.0
TV Comedies		NaN	581.0

1. In Movies :- International Movies > Dramas > Comedies are the most preferred Categories .
2. In TV Shows :- International TV Shows > TV Dramas > TV Comedies are the most preferred Categories .

In [91]:

```
c=pd.crosstab(index=df_final["Year"],columns=df_final["type"],values=df_final["title"]  
c.reset_index(inplace=True)  
c
```

Out[91]:

type	Year	Movie	TV Show
0	2020	1284.0	595.0
1	2019	1424.0	594.0
2	2021	993.0	507.0
3	2018	1237.0	414.0
4	2017	839.0	350.0
5	2016	253.0	179.0
6	2015	56.0	26.0
7	2013	6.0	5.0
8	2014	19.0	5.0
9	2008	1.0	1.0
10	2009	2.0	NaN
11	2010	1.0	NaN
12	2011	13.0	NaN
13	2012	3.0	NaN

In [92]:

```
type_yearly=pd.melt(c,id_vars=["Year"],value_vars=["Movie", "TV Show"])
type_yearly
```

Out[92]:

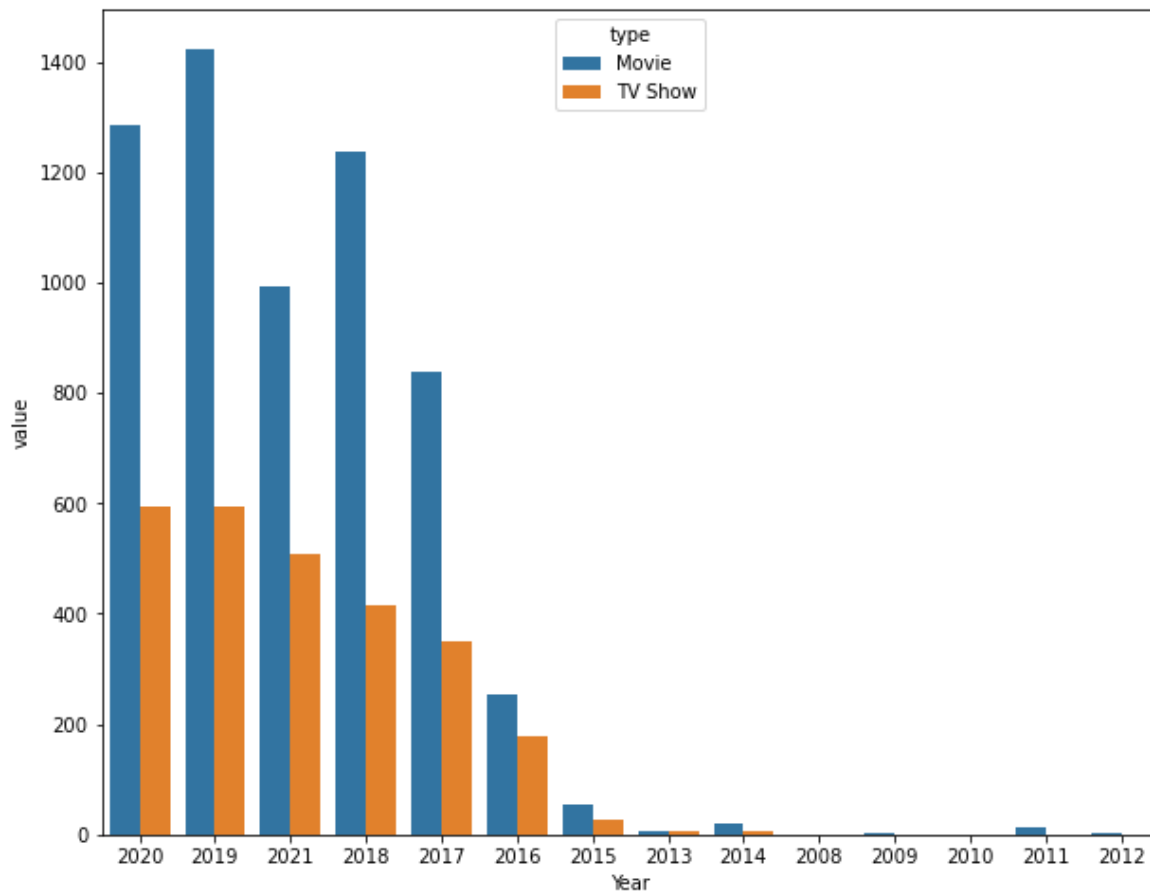
	Year	type	value
0	2020	Movie	1284.0
1	2019	Movie	1424.0
2	2021	Movie	993.0
3	2018	Movie	1237.0
4	2017	Movie	839.0
5	2016	Movie	253.0
6	2015	Movie	56.0
7	2013	Movie	6.0
8	2014	Movie	19.0
9	2008	Movie	1.0
10	2009	Movie	2.0
11	2010	Movie	1.0
12	2011	Movie	13.0
13	2012	Movie	3.0
14	2020	TV Show	595.0
15	2019	TV Show	594.0
16	2021	TV Show	507.0
17	2018	TV Show	414.0
18	2017	TV Show	350.0
19	2016	TV Show	179.0
20	2015	TV Show	26.0
21	2013	TV Show	5.0
22	2014	TV Show	5.0
23	2008	TV Show	1.0
24	2009	TV Show	NaN
25	2010	TV Show	NaN
26	2011	TV Show	NaN
27	2012	TV Show	NaN

In [93]:

```
plt.figure(figsize=(10,8))
sns.barplot(x=type_yearly["Year"],y=type_yearly["value"],hue=type_yearly["type"])
```

Out[93]:

<AxesSubplot:xlabel='Year', ylabel='value'>



1. A **RIISING TREND** is visible over the years in terms of Movies and TV Shows with the peak coming in 2019 .

In [94]:

```
d=pd.crosstab(index=df_final["Country"],columns=df_final["type"],values=df_final["title"],aggfunc='sum')
d.reset_index(inplace=True)
d
```

Out[94]:

	type	Country	Movie	TV Show
0		United States	2940.0	1308.0
1		United Kingdom	556.0	273.0
2		Japan	138.0	200.0
3		South Korea	64.0	171.0
4		Canada	334.0	126.0
...	
123		United States,	1.0	NaN
124		Vatican City	1.0	NaN
125		Venezuela	4.0	NaN
126		Vietnam	7.0	NaN
127		Zimbabwe	3.0	NaN

128 rows × 3 columns

In [95]:

```
top_10_tv=d[:10]
top_10_tv
```

Out[95]:

	type	Country	Movie	TV Show
0		United States	2940.0	1308.0
1		United Kingdom	556.0	273.0
2		Japan	138.0	200.0
3		South Korea	64.0	171.0
4		Canada	334.0	126.0
5		France	318.0	91.0
6		India	1052.0	86.0
7		Taiwan	22.0	72.0
8		Australia	96.0	66.0
9		Spain	176.0	63.0

In [96]:

```
type_country=pd.melt(top_10_tv,id_vars=["Country"],value_vars=["Movie", "TV Show"])
type_country
```

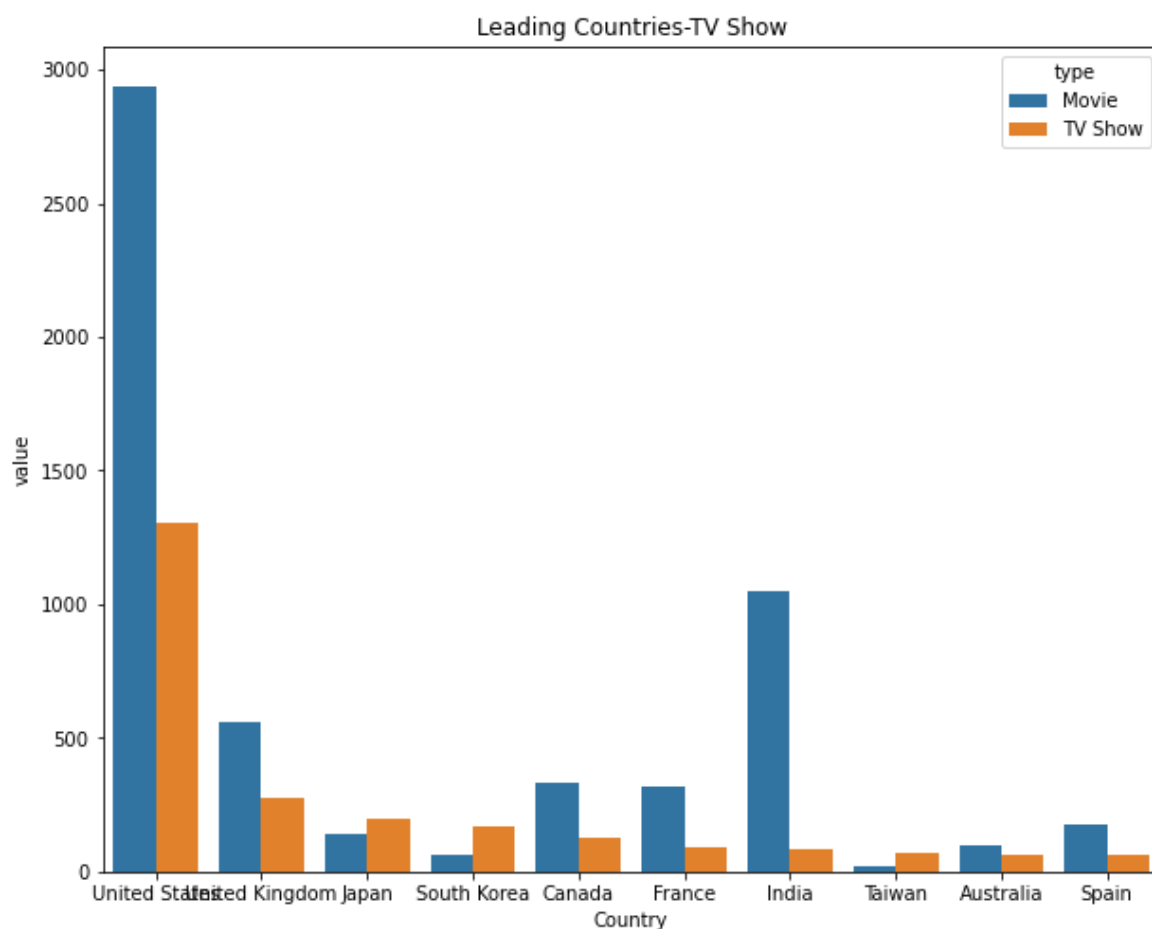
Out[96]:

	Country	type	value
0	United States	Movie	2940.0
1	United Kingdom	Movie	556.0
2	Japan	Movie	138.0
3	South Korea	Movie	64.0
4	Canada	Movie	334.0
5	France	Movie	318.0
6	India	Movie	1052.0
7	Taiwan	Movie	22.0
8	Australia	Movie	96.0
9	Spain	Movie	176.0
10	United States	TV Show	1308.0
11	United Kingdom	TV Show	273.0
12	Japan	TV Show	200.0
13	South Korea	TV Show	171.0
14	Canada	TV Show	126.0
15	France	TV Show	91.0
16	India	TV Show	86.0
17	Taiwan	TV Show	72.0
18	Australia	TV Show	66.0
19	Spain	TV Show	63.0

Top 10 countriesTV Show wise

In [97]:

```
plt.figure(figsize=(10,8))
sns.barplot(x=type_country["Country"],y=type_country["value"],hue=type_country["type"]
plt.title("Leading Countries-TV Show")
plt.show()
```



1. We can see in term of TV Shows :- United States > United kingdom > Japan > South Korea .

In [98]:

```
g=pd.crosstab(index=df_final["Country"],columns=df_final["type"],values=df_final["title"],aggfunc=sum,dropna=False)
g.reset_index(inplace=True)
g
```

Out[98]:

	type	Country	Movie	TV Show
0		United States	2940.0	1308.0
1		India	1052.0	86.0
2		United Kingdom	556.0	273.0
3		Canada	334.0	126.0
4		France	318.0	91.0
...	
123		Somalia	1.0	NaN
124		Azerbaijan	NaN	1.0
125		Belarus	NaN	1.0
126		Cyprus	NaN	1.0
127		Puerto Rico	NaN	1.0

128 rows × 4 columns

In [99]:

```
top_10_movie=g[:10]
top_10_movie
```

Out[99]:

	type	Country	Movie	TV Show
0		United States	2940.0	1308.0
1		India	1052.0	86.0
2		United Kingdom	556.0	273.0
3		Canada	334.0	126.0
4		France	318.0	91.0
5		Germany	187.0	44.0
6		Spain	176.0	63.0
7		Country Not Found	156.0	19.0
8		Japan	138.0	200.0
9		Nigeria	129.0	11.0

In [100]:

```
type_country1=pd.melt(top_10_movie,id_vars=["Country"],value_vars=["Movie", "TV Show"]  
type_country1
```

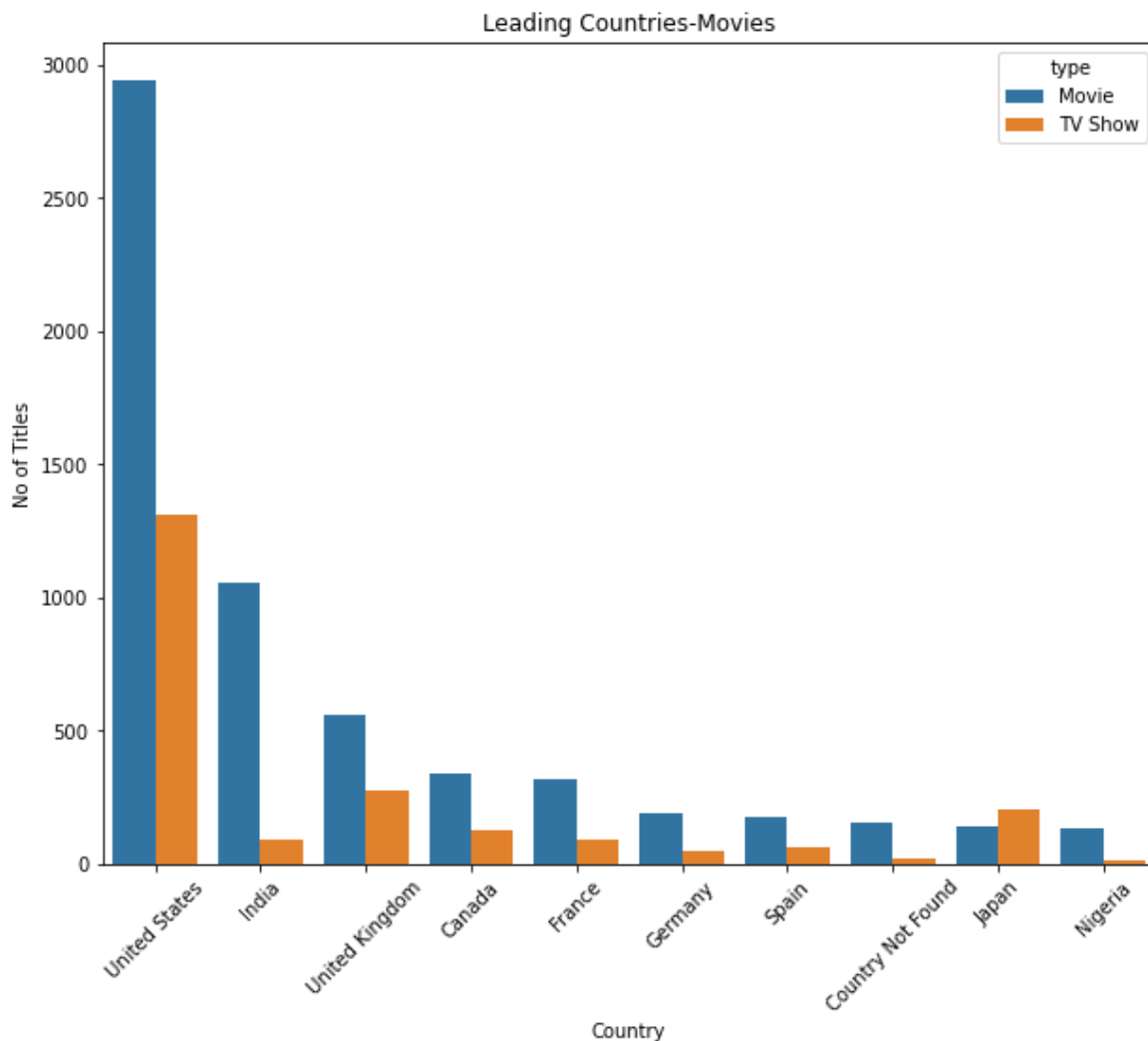
Out[100]:

	Country	type	value
0	United States	Movie	2940.0
1	India	Movie	1052.0
2	United Kingdom	Movie	556.0
3	Canada	Movie	334.0
4	France	Movie	318.0
5	Germany	Movie	187.0
6	Spain	Movie	176.0
7	Country Not Found	Movie	156.0
8	Japan	Movie	138.0
9	Nigeria	Movie	129.0
10	United States	TV Show	1308.0
11	India	TV Show	86.0
12	United Kingdom	TV Show	273.0
13	Canada	TV Show	126.0
14	France	TV Show	91.0
15	Germany	TV Show	44.0
16	Spain	TV Show	63.0
17	Country Not Found	TV Show	19.0
18	Japan	TV Show	200.0
19	Nigeria	TV Show	11.0

Top 10 countries Movie wise

In [101]:

```
plt.figure(figsize=(10,8))
sns.barplot(x=type_country1["Country"],y=type_country1["value"],hue=type_country1["type"])
plt.xticks(rotation=45)
plt.ylabel("No of Titles")
plt.title("Leading Countries-Movies")
plt.show()
```



1. We can see that in terms of Movies , United States > India > United Kingdom > Canada > France
2. Also we can see that in **India** there is a huge gap b/w **TV_shows** and **Movies**

In [102]:

```
df_final["rating"].value_counts()[ :5]
```

Out[102]:

```
TV-MA    73819
TV-14    43925
R         25859
PG-13    16246
TV-PG    14926
Name: rating, dtype: int64
```

In [103]:

```
data_rating=df_final[(df_final["rating"]=="TV-MA")|(df_final["rating"]=="TV-14")|(df_
data_rating
```

Out[103]:

	title	Actors	Directors	Category	Country	show_id	type	date_ad
0	Dick Johnson Is Dead	Actor Unknown	Kirsten Johnson	Documentaries	United States	s1	Movie	2021-0
1	Blood & Water	Ama Qamata	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-0
2	Blood & Water	Ama Qamata	Director Unknown	TV Dramas	South Africa	s2	TV Show	2021-0
3	Blood & Water	Ama Qamata	Director Unknown	TV Mysteries	South Africa	s2	TV Show	2021-0
4	Blood & Water	Khosi Ngema	Director Unknown	International TV Shows	South Africa	s2	TV Show	2021-0
...
201986	Zubaan	Anita Shabdish	Mozez Singh	International Movies	India	s8807	Movie	2019-0
201987	Zubaan	Anita Shabdish	Mozez Singh	Music & Musicals	India	s8807	Movie	2019-0
201988	Zubaan	Chittaranjan Tripathy	Mozez Singh	Dramas	India	s8807	Movie	2019-0
201989	Zubaan	Chittaranjan Tripathy	Mozez Singh	International Movies	India	s8807	Movie	2019-0
201990	Zubaan	Chittaranjan Tripathy	Mozez Singh	Music & Musicals	India	s8807	Movie	2019-0

174775 rows × 13 columns

In [104]:

```
data_rating["rating"].value_counts()
```

Out[104]:

TV-MA 73819
TV-14 43925
R 25859
PG-13 16246
TV-PG 14926
Name: rating, dtype: int64

In-Depth Analysis of Rating for top 5 Countries

In [105]:

```
data_rating_country=data_rating[data_rating["Country"].isin(["United States", "India"])]
data_rating_country
```

Out[105]:

	title	Actors	Directors	Category	Country	show_id	type	date_
0	Dick Johnson Is Dead	Actor Unknown	Kirsten Johnson	Documentaries	United States	s1	Movie	2021
58	Ganglands	Sami Bouajila	Julien Leclercq	Crime TV Shows	France	s3	TV Show	2021
59	Ganglands	Sami Bouajila	Julien Leclercq	International TV Shows	France	s3	TV Show	2021
60	Ganglands	Sami Bouajila	Julien Leclercq	TV Action & Adventure	France	s3	TV Show	2021
61	Ganglands	Tracy Gotoas	Julien Leclercq	Crime TV Shows	France	s3	TV Show	2021
...
201986	Zubaan	Anita Shabdish	Mozez Singh	International Movies	India	s8807	Movie	2019
201987	Zubaan	Anita Shabdish	Mozez Singh	Music & Musicals	India	s8807	Movie	2019
201988	Zubaan	Chittaranjan Tripathy	Mozez Singh	Dramas	India	s8807	Movie	2019
201989	Zubaan	Chittaranjan Tripathy	Mozez Singh	International Movies	India	s8807	Movie	2019
201990	Zubaan	Chittaranjan Tripathy	Mozez Singh	Music & Musicals	India	s8807	Movie	2019

98296 rows × 13 columns



In [106]:

```
data_rating_country1=pd.crosstab(index=data_rating_country["Country"],columns=data_ra
data_rating_country1
```

Out[106]:

	rating	PG-13	R	TV-14	TV-MA	TV-PG
Country						
<hr/>						
United States		438	669	629	1287	372
India		11	6	598	276	148
United Kingdom		84	148	106	259	99
France		36	57	49	172	12
Canada		33	85	52	108	40

In [107]:

```
data_rating_country1.reset_index(inplace=True)
data_rating_country1
```

Out[107]:

	rating	Country	PG-13	R	TV-14	TV-MA	TV-PG
<hr/>							
0		United States	438	669	629	1287	372
1		India	11	6	598	276	148
2		United Kingdom	84	148	106	259	99
3		France	36	57	49	172	12
4		Canada	33	85	52	108	40

In [108]:

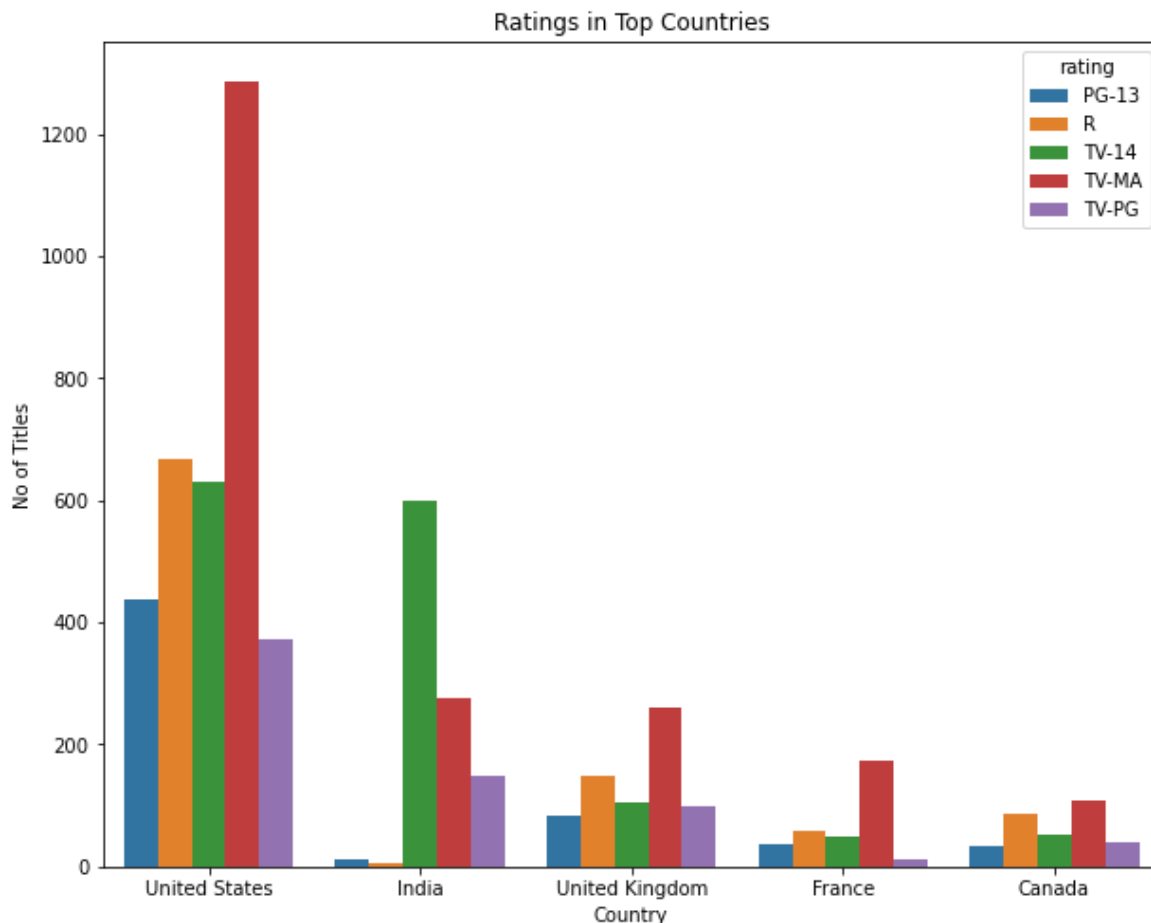
```
data=pd.melt(data_rating_country1,id_vars=["Country"])
data
```

Out[108]:

	Country	rating	value
0	United States	PG-13	438
1	India	PG-13	11
2	United Kingdom	PG-13	84
3	France	PG-13	36
4	Canada	PG-13	33
5	United States	R	669
6	India	R	6
7	United Kingdom	R	148
8	France	R	57
9	Canada	R	85
10	United States	TV-14	629
11	India	TV-14	598
12	United Kingdom	TV-14	106
13	France	TV-14	49
14	Canada	TV-14	52
15	United States	TV-MA	1287
16	India	TV-MA	276
17	United Kingdom	TV-MA	259
18	France	TV-MA	172
19	Canada	TV-MA	108
20	United States	TV-PG	372
21	India	TV-PG	148
22	United Kingdom	TV-PG	99
23	France	TV-PG	12
24	Canada	TV-PG	40

In [109]:

```
plt.figure(figsize=(10,8))
sns.barplot(x=data["Country"],y=data["value"],hue=data["rating"])
plt.ylabel("No of Titles")
plt.title("Ratings in Top Countries")
plt.show()
```



Interesting Observation

1. We can see that TV-MA is the most common rating in all top countries, but India has some different trends.
2. Normally **TV-MA** is followed by **R**.
3. In India Top-Rating is **TV-14** , followed by **TV-MA**

Insights

1. We can clearly see that July is the month when most movies were added
2. Tv-MA is the most common rating
3. United States is leading in both Tv-Shows & Movies
4. Indian Market is different and preference for movies is more here as compared to other countries
5. Also TV-14 is the most prevalent rating here in India
6. On an average there is a rising trend in both no of movies & tv shows

7. Different set of countries leads in TV-Shows as compared to Movies
8. Snowden is the most watched show all over
9. People on an average watch more movies with a duration 80-100 minutes
10. In terms of TV Shows :- 1 season is the leading duration type
11. International Shows , Comedies & Dramas are the leading categories

Recommendations

1. More movies should be produced in India as it is a big market .
2. Better viewability is guaranteed in July and September , so focus on these months
3. The Bottom 10 Categories should be removed from the platform to lower the burden on servers .
4. More short tenure / Duration Movies should be added to the platform
5. The Frequency should be increased on Weekends to improve the viewership