In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
#pd.set_option('display.max_rows', None)
```

In [73]:
```
data = pd.read_csv('netflix.csv')
```

Getting top 5 values from data

In [74]:
```
data.head()
```

Out[74]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentar
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	Internatio TV Shows, Dramas, Myster
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime Sho Internatio TV Shows, Ac
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseri Reality
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	Internatio TV Sho Romantic Shows, TV

Checking shape of data

In [75]:
```
data.shape
```

Out[75]:
```
(8807, 12)
```

Checking data information

This will help us to understand how many records are available in each column

In [76]:
```
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
```

checking for missing data value counts

In [77]:
```
data.isna().sum(axis=0)
```

Out[77]:
```
show_id          0
type             0
title            0
director      2634
cast           825
country        831
date_added      10
release_year     0
rating           4
duration         3
listed_in        0
description      0
dtype: int64
```

In [78]:
```
# statistical data anlysis
```

In [79]:
```
desc_cols=[]
desc_cols_val=[]
```

In [80]:
```
data_desc_numerical = data.describe()
data_desc_numerical
```

Out[80]:

	release_year
count	8807.000000
mean	2014.180198
std	8.819312
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

In [81]:

```
obj_desc = data.describe(include='object')
obj_desc
```

Out[81]:

	show_id	type	title	director	cast	country	date_added	rating	duration	listed_in	desc
count	8807	8807	8807	6173	7982	7976	8797	8803	8804	8807	
unique	8807	2	8807	4528	7692	748	1767	17	220	514	
top	s1	Movie	Dick Johnson Is Dead	Rajiv Chilaka	David Attenborough	United States	January 1, 2020	TV-MA	1 Season	Dramas, International Movies	Para activ abar
freq	1	6131	1	19	19	2818	109	3207	1793	362	

In [82]:

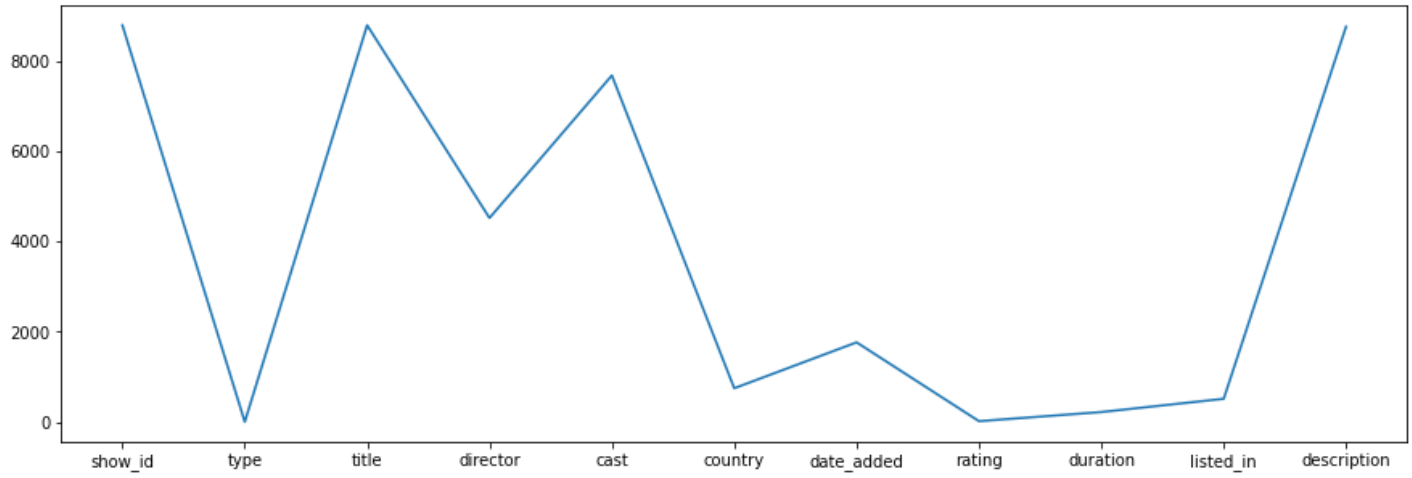
```
for i in obj_desc:
    #print(i,obj_desc[i]['unique'])
    desc_cols.append(i)
    desc_cols_val.append(obj_desc[i]['unique'])
```

ploting unique count for each column

from this plot we can see the difference between counts amoung columns

In [83]:

```
plt.figure(figsize=(15,5))
sns.lineplot(x=desc_cols,y=desc_cols_val)
plt.show()
```



PreProcessing data

In [84]:
```
data.head()
```

Out[84]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentar
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	Internatio TV Shows, Dramas, Myster
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime Sho Internatio TV Shows, Ac
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseri Reality
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	Internatio TV Sho Romantic Shows, TV

In [85]:
```
data['date_added'],data['release_year'] = data['release_year'], data['date_added']
```

In [86]:
```
data.rename({'date_added':'release_year','release_year':'date_added'},axis=1,inplace=True)
```

In [87]:
```
data.head()
```

Out[87]:

	show_id	type	title	director	cast	country	release_year	date_added	rating	duration	listed_
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2020	September 25, 2021	PG-13	90 min	Documentar
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021	September 24, 2021	TV-MA	2 Seasons	Internatio TV Shows, Dramas, Myster
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	2021	September 24, 2021	TV-MA	1 Season	Crime Shov Internatio TV Shows, Ac
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	2021	September 24, 2021	TV-MA	1 Season	Docuseri Reality
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021	September 24, 2021	TV-MA	2 Seasons	Internatio TV Shov Romantic Shows, TV

Below columns we need to unfold for getting deep data anlysis

1. director

2. cast

3. listed_in

4. country

Unfolding director column

In [88]:
```
def get_seperate_cols_values(df,vals):
    constrains = df[vals].apply(lambda x:str(x).strip().split(',')).tolist()
    n_df = pd.DataFrame(constrains,index=data['title'])
    n_df = pd.DataFrame(n_df.stack()).reset_index()
    n_df = n_df[['title',0]]
    n_df.columns = ['title',vals]
    return n_df
```

In [89]:
```
title_wrt_director = get_seperate_cols_values(data,'director')
```

Check for director columns containing comma seperated values

In [90]:
```
title_wrt_director.loc[title_wrt_director['director'].str.contains(',')]
```

Out[90]:

	title	director

In [91]:
```
title_wrt_director.head()
```

Out[91]:

	title	director
0	Dick Johnson Is Dead	Kirsten Johnson
1	Blood & Water	nan
2	Ganglands	Julien Leclercq
3	Jailbirds New Orleans	nan
4	Kota Factory	nan

Unfolding Cast column

In [92]:
```
title_wrt_cast = get_seperate_cols_values(data,'cast')
title_wrt_cast
```

Out[92]:

	title	cast
0	Dick Johnson Is Dead	nan
1	Blood & Water	Ama Qamata
2	Blood & Water	Khosi Ngema
3	Blood & Water	Gail Mabalane
4	Blood & Water	Thabang Molaba
...	...	...
64946	Zubaan	Manish Chaudhary
64947	Zubaan	Meghna Malik

	title	cast
64948	Zubaan	Malkeet Rauni
64949	Zubaan	Anita Shabdish
64950	Zubaan	Chittaranjan Tripathy

64951 rows × 2 columns

Unfolding listed_in column

In [93]:
```
title_wrt_listed_in = get_seperate_cols_values(data,'listed_in')
title_wrt_listed_in
```

Out[93]:

	title	listed_in
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

In [94]:
```
np.any(title_wrt_listed_in['listed_in'].str.contains(','))
```

Out[94]:
```
False
```

Unfolding country column

In [95]:
```
title_wrt_country = get_seperate_cols_values(data,'country')
title_wrt_country
```

Out[95]:

	title	country
0	Dick Johnson Is Dead	United States
1	Blood & Water	South Africa
2	Ganglands	nan
3	Jailbirds New Orleans	nan

	title	country
4	Kota Factory	India
...	...	...
10845	Zodiac	United States
10846	Zombie Dumb	nan
10847	Zombieland	United States
10848	Zoom	United States
10849	Zubaan	India

10850 rows × 2 columns

lets join these all unfolded data

merging director and cast

In [96]:
```
merged_title_director = title_wrt_director.merge(title_wrt_cast,on='title')
```

In [97]:
```
merged_title_director.head()
```

Out[97]:

	title	director	cast
0	Dick Johnson Is Dead	Kirsten Johnson	nan
1	Blood & Water	nan	Ama Qamata
2	Blood & Water	nan	Khosi Ngema
3	Blood & Water	nan	Gail Mabalane
4	Blood & Water	nan	Thabang Molaba

merging listed_in and country

In [98]:
```
merged_listed_in_and_country = title_wrt_listed_in.merge(title_wrt_country,on='title')
```

In [99]:
```
merged_listed_in_and_country.head()
```

Out[99]:

	title	listed_in	country
0	Dick Johnson Is Dead	Documentaries	United States
1	Blood & Water	International TV Shows	South Africa
2	Blood & Water	TV Dramas	South Africa
3	Blood & Water	TV Mysteries	South Africa
4	Ganglands	Crime TV Shows	nan

merging merged_listed_in_and_country and merged_title_director

In [100…
```
final_merged_df = merged_listed_in_and_country.merge(merged_title_director,on='title')
```

In [101…
```
data = data.merge(final_merged_df,on='title',how='inner')
```

In [102…
```
data.head()
```

Out[102…

	show_id	type	title	director_x	cast_x	country_x	release_year	date_added	rating	duration	listed_
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2020	September 25, 2021	PG-13	90 min	Document
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021	September 24, 2021	TV-MA	2 Seasons	Internat TV Show Drama Myst
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021	September 24, 2021	TV-MA	2 Seasons	Internat TV Show Drama Myst
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021	September 24, 2021	TV-MA	2 Seasons	Internat TV Show Drama Myst
4	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021	September 24, 2021	TV-MA	2 Seasons	Internat TV Show Drama Myst

dropping unnecessary columns and renaming some columns

In [103…
```
data.drop(['director_x','cast_x','country_x','listed_in_x'],axis=1,inplace=True)
```

In [104…
```
data.rename({'listed_in_y':'listed_in','country_y':'country','director_y':'director','cast
```

In [105…
```
data.columns
```

Out[105…
```
Index(['show_id', 'type', 'title', 'release_year', 'date_added', 'rating',
       'duration', 'description', 'listed_in', 'country', 'director', 'cast'],
      dtype='object')
```

In [106…
```
data.drop('description',axis=1,inplace=True)
```

Replacing txt 'nan' with np.nan values for better data preprocessing

In [107…
```
data.loc[data['cast'].isin(['nan'])]
```

Out[107…

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director
0	s1	Movie	Dick Johnson Is Dead	2020	September 25, 2021	PG-13	90 min	Documentaries	United States	Kirsten Johnson
85	s4	TV Show	Jailbirds New Orleans	2021	September 24, 2021	TV-MA	1 Season	Docuseries	nan	nan
86	s4	TV Show	Jailbirds New Orleans	2021	September 24, 2021	TV-MA	1 Season	Reality TV	nan	nan
353	s11	TV Show	Vendetta: Truth, Lies and The Mafia	2021	September 24, 2021	TV-MA	1 Season	Crime TV Shows	nan	nan
354	s11	TV Show	Vendetta: Truth, Lies and The Mafia	2021	September 24, 2021	TV-MA	1 Season	Docuseries	nan	nan
…	…	…	…	…	…	…	…	…	…	…
200779	s8757	Movie	Woodstock	2019	August 13, 2019	TV-MA	97 min	Music & Musicals	United States	Barak Goodman
200898	s8764	Movie	WWII: Report from the Aleutians	1943	March 31, 2017	TV-PG	45 min	Documentaries	United States	John Huston
202006	s8804	TV Show	Zombie Dumb	2018	July 1, 2019	TV-Y7	2 Seasons	Kids' TV	nan	nan
202007	s8804	TV Show	Zombie Dumb	2018	July 1, 2019	TV-Y7	2 Seasons	Korean TV Shows	nan	nan
202008	s8804	TV Show	Zombie Dumb	2018	July 1, 2019	TV-Y7	2 Seasons	TV Comedies	nan	nan

2149 rows × 11 columns

In [108…
```
data['cast'].replace('nan',np.nan,inplace=True)
```

In [109…

```
data.loc[data['director'].isin(['nan'])]
```

Out[109…

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	c
1	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	nan	A Qam
2	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	nan	Kh Nge
3	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	nan	Mabal
4	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	nan	Thab Mol
5	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	nan	Di Windvc
…	…	…	…	…	…	…	…	…	…	…	
201938	s8801	TV Show	Zindagi Gulzar Hai	2012	December 15, 2016	TV-PG	1 Season	TV Dramas	Pakistan	nan	Ja She
201939	s8801	TV Show	Zindagi Gulzar Hai	2012	December 15, 2016	TV-PG	1 Season	TV Dramas	Pakistan	nan	H Khav Ba
202006	s8804	TV Show	Zombie Dumb	2018	July 1, 2019	TV-Y7	2 Seasons	Kids' TV	nan	nan	N
202007	s8804	TV Show	Zombie Dumb	2018	July 1, 2019	TV-Y7	2 Seasons	Korean TV Shows	nan	nan	N
202008	s8804	TV Show	Zombie Dumb	2018	July 1, 2019	TV-Y7	2 Seasons	TV Comedies	nan	nan	N

50643 rows × 11 columns

In [110…

```
data['director'].replace('nan',np.nan,inplace=True)
```

In [111…

```
data.loc[data['country'].isin(['nan'])]
```

Out[111…

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	cas
58	s3	TV Show	Ganglands	2021	September 24, 2021	TV-MA	1 Season	Crime TV Shows	nan	Julien Leclercq	Sam Bouajil
59	s3	TV Show	Ganglands	2021	September 24, 2021	TV-MA	1 Season	Crime TV Shows	nan	Julien Leclercq	Trac Gotoa
60	s3	TV Show	Ganglands	2021	September 24, 2021	TV-MA	1 Season	Crime TV Shows	nan	Julien Leclercq	Samue Jou
61	s3	TV Show	Ganglands	2021	September 24, 2021	TV-MA	1 Season	Crime TV Shows	nan	Julien Leclercq	Nabih Akka

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	cas
62	s3	TV Show	Ganglands	2021	September 24, 2021	TV-MA	1 Season	Crime TV Shows	nan	Julien Leclercq	Sofi Lesaffr
...	...	...	...	...	...	...	...	...	...	...	.
201498	s8786	TV Show	YOM	2016	June 7, 2018	TV-Y7	1 Season	Kids' TV	nan	NaN	Mayu Vya
201499	s8786	TV Show	YOM	2016	June 7, 2018	TV-Y7	1 Season	Kids' TV	nan	NaN	Ketar Kav
202006	s8804	TV Show	Zombie Dumb	2018	July 1, 2019	TV-Y7	2 Seasons	Kids' TV	nan	NaN	NaN
202007	s8804	TV Show	Zombie Dumb	2018	July 1, 2019	TV-Y7	2 Seasons	Korean TV Shows	nan	NaN	NaN
202008	s8804	TV Show	Zombie Dumb	2018	July 1, 2019	TV-Y7	2 Seasons	TV Comedies	nan	NaN	NaN

11897 rows × 11 columns

In [112…
```
data['country'].replace('nan',np.nan,inplace=True)
```

Final NA Count in data

In [113…
```
data.isna().sum()
```

Out[113…
```
show_id          0
type             0
title            0
release_year     0
date_added     158
rating          67
duration         3
listed_in        0
country      11897
director     50643
cast          2149
dtype: int64
```

Fixing NAN values for data

In [114…
```
data
```

Out[114…

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director
0	s1	Movie	Dick Johnson Is Dead	2020	September 25, 2021	PG-13	90 min	Documentaries	United States	Kirsten Johnson
1	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	NaN
2	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	NaN

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	
3	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	NaN	N
4	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	NaN	
...	...	...	...	...	...	...	...	...	...	...	
202060	s8807	Movie	Zubaan	2015	March 2, 2019	TV-14	111 min	Music & Musicals	India	Mozez Singh	Ch
202061	s8807	Movie	Zubaan	2015	March 2, 2019	TV-14	111 min	Music & Musicals	India	Mozez Singh	
202062	s8807	Movie	Zubaan	2015	March 2, 2019	TV-14	111 min	Music & Musicals	India	Mozez Singh	
202063	s8807	Movie	Zubaan	2015	March 2, 2019	TV-14	111 min	Music & Musicals	India	Mozez Singh	
202064	s8807	Movie	Zubaan	2015	March 2, 2019	TV-14	111 min	Music & Musicals	India	Mozez Singh	Chi

202065 rows × 11 columns

Creating group on the basis of listed_in to fix country column NAN values

Assumption : checking for most occuring genre in specific country this will tell which genre is popular in each country

In [115…
```
grp_listed_in = data.groupby('listed_in').agg(pd.Series.mode).reset_index()
```

In [116…
```
listed_in_vals = grp_listed_in[['listed_in','country']]
```

In [117…
```
listed_in_vals.loc[listed_in_vals['listed_in']=='Reality TV']['country']
```

Out[117…
```
58    United States
Name: country, dtype: object
```

In [118…
```
def isNaN(string):
    return string != string
```

In [119…
```
df_country_lis_in = data[['listed_in','country']]
```

In [120…
```
x=df_country_lis_in['listed_in']
y= df_country_lis_in['country']
for i,j in zip(range(len(x)),range(len(y))):
    if(isNaN(y[j])==True):
        show_name = x[i]
```

```
        #print(show_name,type(show_name))
        #print(x[i],y[i],listed_in_vals.loc[listed_in_vals['listed_in']==show_name]['coun
        y[i] = listed_in_vals.loc[listed_in_vals['listed_in']==show_name]['country'].value
```

```
C:\Users\PC-DELL\AppData\Local\Temp\ipykernel_23400\1875247350.py:8: SettingWithCopyWarnin
g:
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_gu
ide/indexing.html#returning-a-view-versus-a-copy
  y[i] = listed_in_vals.loc[listed_in_vals['listed_in']==show_name]['country'].values[0]
```

In [121…
```
data['country'] = y
```

In [122…
```
data['country'].isna().sum()
```

Out[122…
```
0
```

In [123…
```
data.isna().sum()
```

Out[123…
```
show_id             0
type                0
title               0
release_year        0
date_added        158
rating             67
duration            3
listed_in           0
country             0
director        50643
cast             2149
dtype: int64
```

Country NAN Values are fixed Now

In [ ]:
```

```

In [124…
```
data.head()
```

Out[124…

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

In [125…
```
data.shape
```

Out[125…
```
(202065, 11)
```

Fixing NAN Values for Director column

Assumption : assuming that director belong to certain country and genre.

So replacing NAN value with most occuring value of country and genre

In [126…
```
grp_listed_in = data.groupby(['listed_in','country']).agg(pd.Series.mode).reset_index()
```

In [127…
```
director_values = grp_listed_in[['listed_in','country','director']]
```

In [128…
```
director_values.shape
```

Out[128…
```
(2440, 3)
```

In [129…
```
director_values
```

Out[129…

	listed_in	country	director
0	Anime Features	Japan	Steven Yamamoto
1	Anime Features	United States	Mamoru Hosoda
2	Anime Features	China	Zhao Ji
3	Anime Features	Japan	Toshiya Shinohara
4	Anime Features	United States	Steven Yamamoto
...	...	...	...
2435	Thrillers	France	John Madden
2436	Thrillers	Singapore	David Chirchirillo
2437	Thrillers	Spain	Maria Pulera
2438	Thrillers	United Kingdom	Sydney Pollack
2439	Thrillers	United States	Ron Howard

2440 rows × 3 columns

In [130…
```
import numpy
```

In [131…
```
temp=[]
dirt = director_values['director']
```

```
liis = director_values['listed_in']
cont = director_values['country']

for i,j,k in zip(director_values['director'],director_values['listed_in'],director_values
    if(type(i)==numpy.ndarray):
        vals = i
        genre = j
        country = k
        #print(vals,genre,country)
        if(len(vals)==0):
            tval = []
            tval.append(genre)
            tval.append(country)
            tval.append('anounimous')
            temp.append(tval)
        for lis in vals:
            t=[]
            t.append(genre)
            t.append(country)
            t.append(lis)
            temp.append(t)
            #print(t)
```

In [132…

```
tdf = pd.DataFrame(temp,columns=['listed_in', 'country','director'])
```

In [133…

```
director_values.append(tdf,ignore_index=True)
```

```
C:\Users\PC-DELL\AppData\Local\Temp\ipykernel_23400\1009010427.py:1: FutureWarning: The fr
ame.append method is deprecated and will be removed from pandas in a future version. Use p
andas.concat instead.
  director_values.append(tdf,ignore_index=True)
```

Out[133…

	listed_in	country	director
0	Anime Features	Japan	Steven Yamamoto
1	Anime Features	United States	Mamoru Hosoda
2	Anime Features	China	Zhao Ji
3	Anime Features	Japan	Toshiya Shinohara
4	Anime Features	United States	Steven Yamamoto
...	...	...	...
4047	TV Shows	United States	Joe Berlinger
4048	Thrillers	Canada	John Madden
4049	Thrillers	Canada	Lee Tamahori
4050	Thrillers	Germany	Lee Tamahori
4051	Thrillers	Germany	Sydney Pollack

4052 rows × 3 columns

In [134…

```
d_indx = director_values.loc[director_values['director'].apply(lambda x : type(x)==numpy.r
```

In [135…

```
d_indx
```

Out[135…
```
Int64Index([   5,   10,   33,   37,   39,   40,   45,   46,   54,   58,
            ...
            2411, 2413, 2414, 2416, 2417, 2418, 2419, 2420, 2422, 2424],
           dtype='int64', length=934)
```

In [136…
```
director_values.drop(d_indx,inplace=True,axis=0)
```

```
C:\Users\PC-DELL\AppData\Local\Temp\ipykernel_23400\1259501998.py:1: SettingWithCopyWarnin
g:
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_gu
ide/indexing.html#returning-a-view-versus-a-copy
  director_values.drop(d_indx,inplace=True,axis=0)
```

In [137…
```
director_values = director_values.append(tdf,ignore_index=True)
```

```
C:\Users\PC-DELL\AppData\Local\Temp\ipykernel_23400\3749041769.py:1: FutureWarning: The fr
ame.append method is deprecated and will be removed from pandas in a future version. Use p
andas.concat instead.
  director_values = director_values.append(tdf,ignore_index=True)
```

In [ ]:
```

```

In [138…
```
data = data.merge(director_values,left_on=['listed_in','country'],right_on=['listed_in','c
```

In [139…
```
data.drop('director_x',axis=1,inplace=True)
```

In [140…
```
data.rename({'director_y':'director'},axis=1,inplace=True)
```

In [141…
```
data.isna().sum()
```

Out[141…
```
show_id          0
type             0
title            0
release_year     0
date_added     170
rating          67
duration         3
listed_in        0
country          0
cast          3913
director         0
dtype: int64
```

Director column NAN Value is Fixed

In [ ]:
```

```

In [142…
```
data.head()
```

Out[142…

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	cast	di

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	cast	di
0	s1	Movie	Dick Johnson Is Dead	2020	September 25, 2021	PG-13	90 min	Documentaries	United States	NaN	
1	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	2020	September 22, 2021	TV-MA	67 min	Documentaries	United States	NaN	
2	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	2020	September 22, 2021	TV-MA	67 min	Documentaries	United States	NaN	
3	s46	Movie	My Heroes Were Cowboys	2021	September 16, 2021	PG	23 min	Documentaries	United States	NaN	
4	s69	Movie	Schumacher	2021	September 15, 2021	TV-14	113 min	Documentaries	United States	Michael Schumacher	

Fixing the NAN values for cast columns

In [143...]
```
data
```

Out[143...]

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	c
0	s1	Movie	Dick Johnson Is Dead	2020	September 25, 2021	PG-13	90 min	Documentaries	United States	N
1	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	2020	September 22, 2021	TV-MA	67 min	Documentaries	United States	N
2	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	2020	September 22, 2021	TV-MA	67 min	Documentaries	United States	N
3	s46	Movie	My Heroes Were Cowboys	2021	September 16, 2021	PG	23 min	Documentaries	United States	N
4	s69	Movie	Schumacher	2021	September 15, 2021	TV-14	113 min	Documentaries	United States	Mich Schumac
...	...	...	...	...	...	...	...	...	...	
258947	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	Christ Ma Caba

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	c
258948	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	Christop Sn
258949	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	M Mitteln
258950	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	Reba B
258951	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	Kyle Heb

258952 rows × 11 columns

In [ ]:

In [ ]:

Creating Group on director,country,listed_in

Assumption : Assuming that every cast is belonging to certain director ,country,listed_in(genre) so replacing the nan value with most occuring value

In [145…
```
dt_grp_for_cast = data.groupby(['director','country','listed_in']).agg(pd.Series.mode).res
```

In [146…
```
dt_grp_for_cast = dt_grp_for_cast[['director','country','listed_in','cast']]
```

In [147…
```
dt_grp_for_cast.head()
```

Out[147…

	director	country	listed_in	cast
0	Aaron Woolf	Venezuela	Documentaries	[]
1	Abdullah Al Noor	Bangladesh	Dramas	[Allen Shubhro, Intekhab Dinar, Iresh Zaker...
2	Abdullah Al Noor	Bangladesh	Independent Movies	[Allen Shubhro, Intekhab Dinar, Iresh Zaker...
3	Abdullah Al Noor	Bangladesh	Comedies	[Allen Shubhro, Intekhab Dinar, Iresh Zaker...
4	Abhishek Chaubey	South Korea	International TV Shows	[Abdurrahman Arif, Arawinda Kirana, Bae Doo...

Unfolding the cast as per row value

In [148…
```
dir_val, cntry, lisin, cst=dt_grp_for_cast['director'], dt_grp_for_cast['country'], dt_grp
```

In [149…
```
cst_lis=[]
for i,j,k,l in zip(dir_val, cntry, lisin, cst):
```

```
    if(type(l)==numpy.ndarray):
        if(len(l)==0):
            tmp=[]
            tmp.append(i)
            tmp.append(j)
            tmp.append(k)
            tmp.append('anounimous')
            cst_lis.append(tmp)
        if(len(l)>0):
            for h in l:
                t=[]
                t.append(i)
                t.append(j)
                t.append(k)
                t.append(h)
                cst_lis.append(t)
                #print(i,j,k,h)
            #print(i,j,k,l)
```

In [150…
```
cst_fixed = pd.DataFrame(cst_lis,columns=['director','country','listed_in','cast'])
```

In [151…
```
data = data.merge(cst_fixed,left_on=['director','country','listed_in'],right_on=['director
```

In [152…
```
data.drop('cast_x',axis=1,inplace=True)
```

In [153…
```
data.rename({'cast_y':'cast'},axis=1,inplace=True)
```

In [154…
```
data.isna().sum()
```

Out[154…
```
show_id          0
type             0
title            0
release_year     0
date_added     807
rating          34
duration        30
listed_in        0
country          0
director         0
cast             0
dtype: int64
```

In [155…
```
data.head()
```

Out[155…

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	cast
0	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Arno Greeff
1	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Chi Mhende
2	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Cindy Mahlangu

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	cast
3	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Dillon Windvogel
4	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Duane Williams

In [156…
```
data['rating'].unique()
```

Out[156…
```
array(['TV-MA', 'TV-14', 'TV-Y7', 'TV-PG', 'TV-G', 'NR', 'R', 'PG',
       'PG-13', 'NC-17', 'TV-Y', 'G', 'TV-Y7-FV', 'UR', '74 min',
       '84 min', '66 min', nan], dtype=object)
```

In [ ]:
```
```

In [157…
```
data.isna().sum()
```

Out[157…
```
show_id          0
type             0
title            0
release_year     0
date_added     807
rating          34
duration        30
listed_in        0
country          0
director         0
cast             0
dtype: int64
```

Fixing NAN Values for duration

Checking NAN value of duration column

Assumption : Assuming that duration depends on type column so filling nan value with most occuring value

In [158…
```
data.groupby('type').agg(pd.Series.mode)
```

Out[158…

type	show_id	title	release_year	date_added	rating	duration	listed_in	country	director	
Movie	s3125	Sincerely Yours, Dhaka	2018	December 16, 2019	TV-MA	106 min	Independent Movies	United Kingdom	Lars von Trier	Anupam

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	
TV Show	s298	Navarasa	2020	August 6, 2021	TV-MA	1 Season	International TV Shows	United States	anounimous	Abdurrah Arif, Araw Kir Brig

In [159…
```
data['duration'].fillna('106 min',axis=0,inplace=True)
```

duration column nan values is fixed

In [160…
```
data.isna().sum()
```

Out[160…
```
show_id           0
type              0
title             0
release_year      0
date_added      807
rating           34
duration          0
listed_in         0
country           0
director          0
cast              0
dtype: int64
```

Fixing NAN value issue for Rating column

Assumption : Filling nan value for rating column with most occuring value in rating column

In [161…
```
data['rating'].mode()[0]
```

Out[161…
```
'TV-MA'
```

In [162…
```
data['rating'].fillna(data['rating'].mode()[0],axis=0,inplace=True)
```

NAN value issue fixed for rating column

In [164…
```
data.isna().sum()
```

Out[164…
```
show_id           0
type              0
title             0
release_year      0
date_added      807
rating            0
duration          0
```

```
listed_in         0
country           0
director          0
cast              0
dtype: int64
```

Fixing NAN value issue for date added column

Assumption :Assumpting that movie/tv show had released late so replacing nan values with most occuring value in date added columns

In [175…
```
data['date_added'].mode()
```

Out[175…
```
0    January 1, 2020
Name: date_added, dtype: object
```

In [166…
```
data['date_added'].fillna(data['date_added'].mode()[0],axis=0,inplace=True)
```

In [167…
```
data.isna().sum()
```

Out[167…
```
show_id         0
type            0
title           0
release_year    0
date_added      0
rating          0
duration        0
listed_in       0
country         0
director        0
cast            0
dtype: int64
```

date_added column nan values are issue fixed

In [ ]:
```

```

In [176…
```
data.isna().sum()
```

Out[176…
```
show_id         0
type            0
title           0
release_year    0
date_added      0
rating          0
duration        0
listed_in       0
country         0
director        0
cast            0
dtype: int64
```

In [171…
```
data.head()
```

Out[171...

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	cast
0	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Arno Greeff
1	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Chi Mhende
2	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Cindy Mahlangu
3	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Dillon Windvogel
4	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Duane Williams

Preprocessing Completed

In [178...

```
data
```

Out[178...

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	
0	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	
1	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	
2	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	M
3	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	W
4	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	
...	...	...	...	...	...	...	...	...	...	...	...
1985046	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous	
1985047	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous	
1985048	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous	M
1985049	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous	R

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director
1985050	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous

1985051 rows × 11 columns

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
#pd.set_option('display.max_rows', None)
```

Data Analysis

In [195…
```
data = pd.read_csv('final_netfilx_cleaned.csv')
```

In [196…
```
data
```

Out[196…

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director
0	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous
1	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous
2	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous
3	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous
4	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous
...	...	...	...	...	...	...	...	...	...	...
1985046	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous
1985047	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous
1985048	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous
1985049	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous
1985050	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous

1985051 rows × 11 columns

Most popular type i.e (Movie/TV Show) among all country

In [4]:
```
grp_by_country = data.groupby('country').agg(pd.Series.mode).reset_index()
```

In [5]:
```
grp_by_country.head()
```

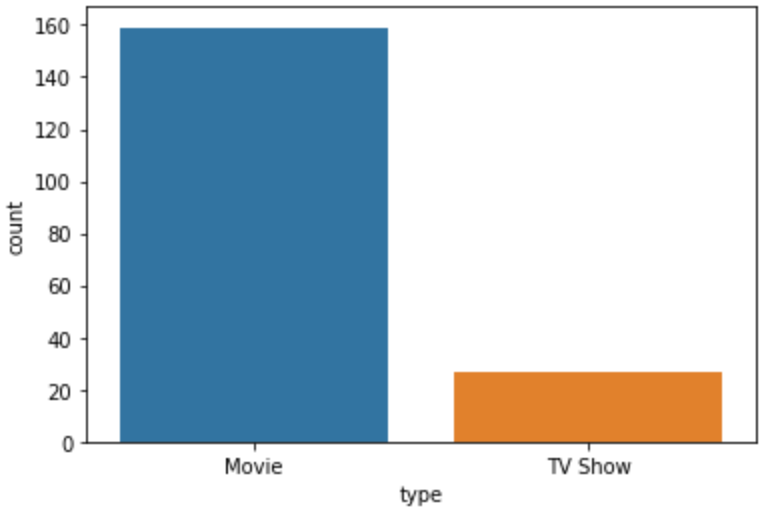
Out[5]:

	country	show_id	type	title	release_year	date_added	rating	duration	listed_in	director	
0	Albania	s4879	Movie	Forgive Us Our Debts	2018	May 4, 2018	TV-MA	105 min	[International Movies, Dramas]	Antonio Morabito	Sa
1	Algeria	s2366	Movie	Alexandria ... Why?	1979	June 18, 2020	TV-MA	131 min	International Movies	Youssef Chahine	M
2	Angola	s2088	Movie	Santana	2020	August 28, 2020	TV-MA	107 min	[International Movies, Action & Adventure]	[Maradona Dias Dos Santos, Chris Roland]	Saty Hak
3	Argentina	s4065	Movie	The Last Runway	2018	March 1, 2019	TV-MA	107 min	Dramas	Álvaro Brechner	A G
4	Armenia	s6234	Movie	Barbecue	2017	August 15, 2017	TV-MA	101 min	[International Movies, Documentaries]	Matthew Salleh	anc

In [6]:
```
import numpy
idx = grp_by_country.loc[grp_by_country['type'].apply(lambda x : type(x)==numpy.ndarray)].
```

In [7]:
```
grp_by_country.drop(idx,axis=0,inplace=True)
```

In [8]:
```
sns.countplot(data=grp_by_country,x='type')
plt.show()
```



Top 10 genre among all country

In [9]:
```
popular_genre=[]
for i,j in zip(grp_by_country['country'],grp_by_country['listed_in']):
    if(type(j)==numpy.ndarray):
        val= j
        for v in val:
            tmp=[]
            tmp.append(i)
            tmp.append(v)
            popular_genre.append(tmp)
    else:
        t=[]
        t.append(i)
        t.append(j)
        popular_genre.append(t)
```

In [10]:
```
popular_genre = pd.DataFrame(popular_genre,columns=['country','genre'])
```

In [11]:
```
popular_genre
```

Out[11]:

	country	genre
0	Albania	International Movies
1	Albania	Dramas
2	Algeria	International Movies
3	Angola	International Movies
4	Angola	Action & Adventure
...	...	...
266	West Germany	International Movies
267	West Germany	Documentaries
268	Zimbabwe	International Movies
269	Zimbabwe	Romantic Movies
270	Zimbabwe	Comedies

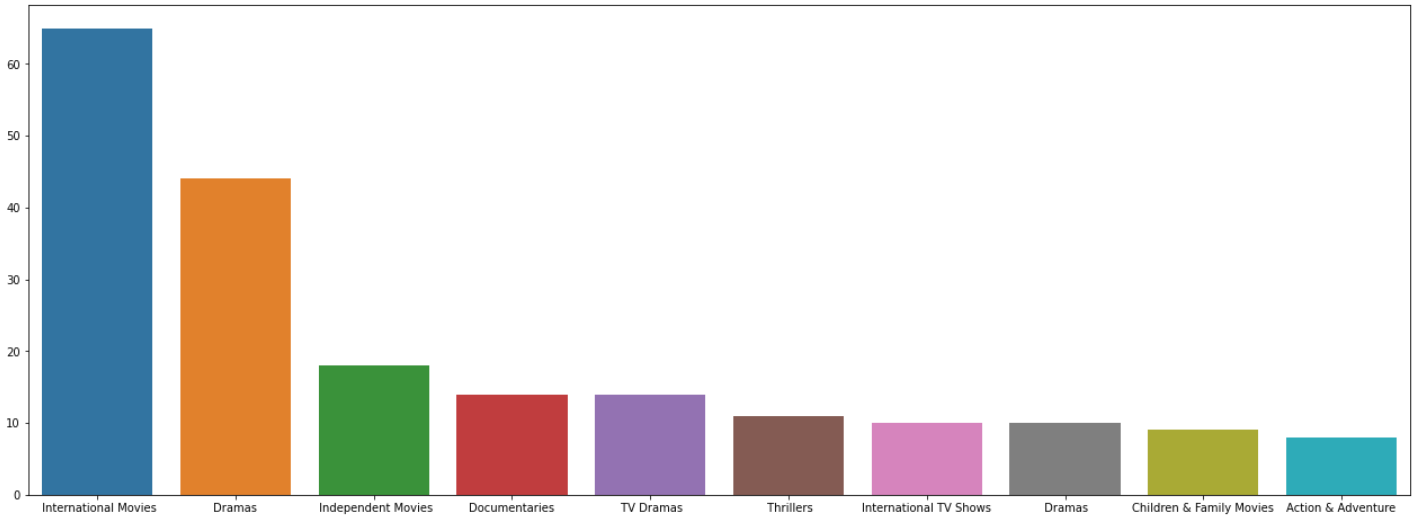
271 rows × 2 columns

In [12]:
```
most_popular_genre_among_all_country = popular_genre['genre'].value_counts()[:10]
```

In [13]:
```
most_popular_genre_among_all_country.index.values
```

Out[13]:
```
array([' International Movies', 'Dramas', ' Independent Movies',
       'Documentaries', ' TV Dramas', ' Thrillers',
       'International TV Shows', ' Dramas', 'Children & Family Movies',
       'Action & Adventure'], dtype=object)
```

In [14]:
```
plt.figure(figsize=(22,8))
sns.barplot(x=most_popular_genre_among_all_country.index.values,y=most_popular_genre_among
plt.show()
```



In [ ]:
```
```

Year wise analysis for some genre

In [15]:
```
data.head()
```

Out[15]:

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	cast
0	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Arno Greeff
1	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Chi Mhende
2	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Cindy Mahlangu
3	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Dillon Windvogel

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	cast
4	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Duane Williams

In [16]:
```
df = pd.DataFrame(data.groupby('listed_in')['release_year'].value_counts()[:50])
```

In [17]:
```
df
```

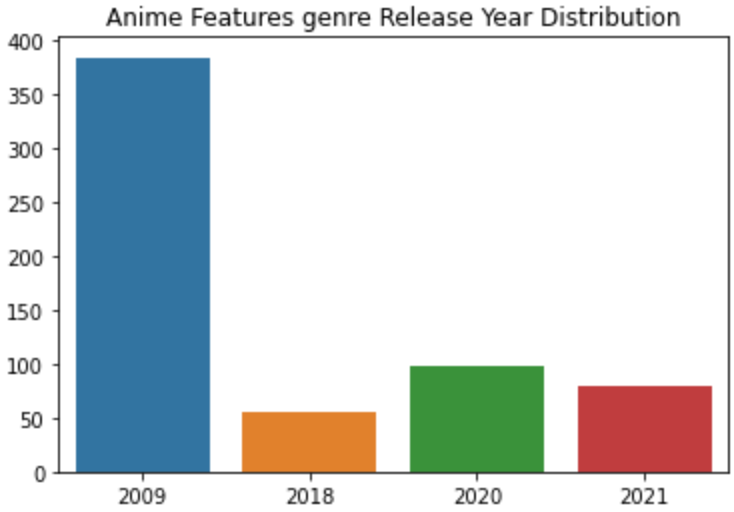
Out[17]:

		release_year
listed_in	**release_year**	
Anime Features	**2009**	384
	2020	98
	2021	80
	2018	56
Children & Family Movies	**2018**	429
	1986	330
	2020	330
	2008	200
	2016	83
	2021	64
Classic & Cult TV	**1977**	288
	2015	252
	2018	252
	1974	216
	2013	144
	2020	144
Classic Movies	**1977**	476
	1986	340
	1990	340
	1968	271
	1971	200
	1975	180
	1979	145
	1973	121
	1961	100
	1963	100
Comedies	**2019**	6244
	2020	5986

	release_year
listed_in	release_year
2017	4300
2009	3665
2018	3411
2000	2922
2015	2869
2008	2618
2016	2551
2014	2179
2011	1967
2021	1664
2004	1575
2002	1472
2013	1442
2010	1419
2012	837
2007	770
1999	684
2001	652
2006	557
2005	336
2003	276
1955	246

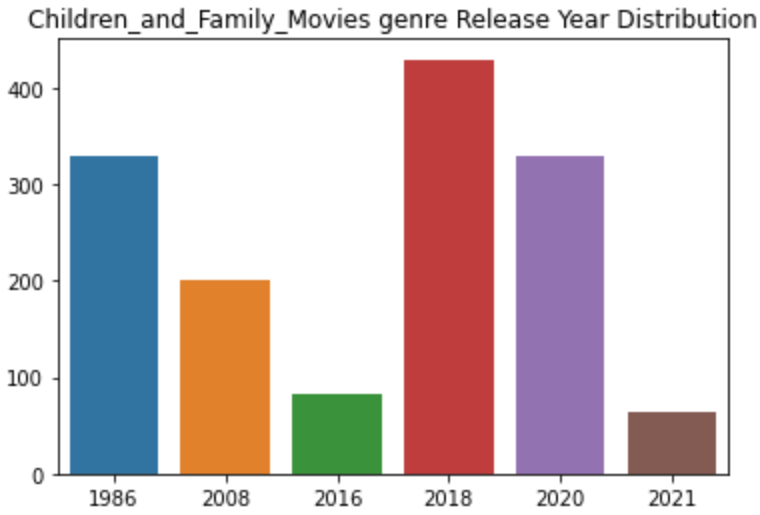
In [18]:
```
anime = df['release_year'][' Anime Features']
```

In [19]:
```
plt.title('Anime Features genre Release Year Distribution')
sns.barplot(x=anime.index.values,y=anime.values)
plt.show()
```



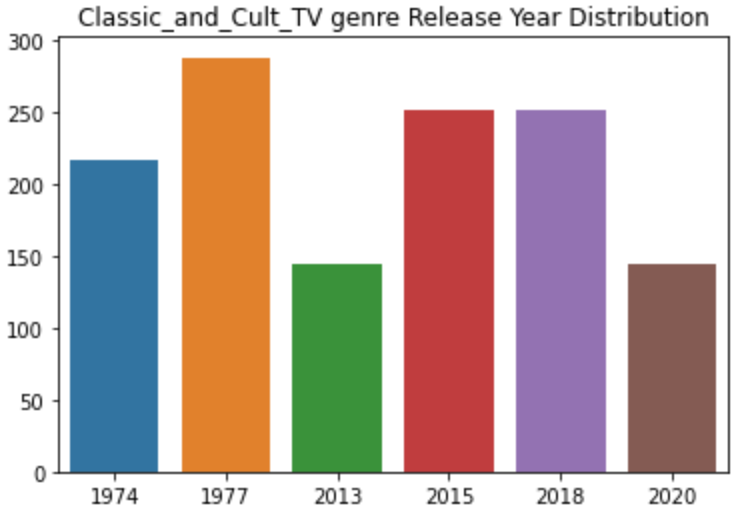
In [20]:
```
Children_and_Family_Movies = df['release_year'][' Children & Family Movies']
```

In [21]:
```
plt.title('Children_and_Family_Movies genre Release Year Distribution')
sns.barplot(x=Children_and_Family_Movies.index.values,y=Children_and_Family_Movies.values)
plt.show()
```



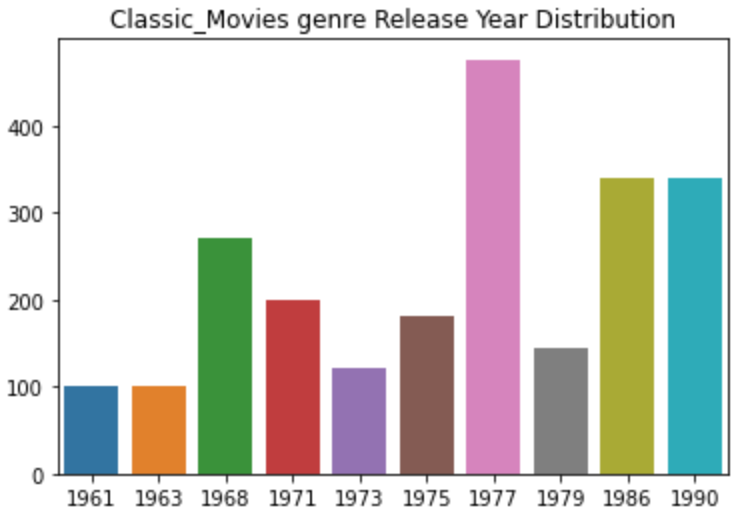
In [22]:
```
Classic_and_Cult_TV = df['release_year'][' Classic & Cult TV']
```

In [23]:
```
plt.title('Classic_and_Cult_TV genre Release Year Distribution')
sns.barplot(x=Classic_and_Cult_TV.index.values,y=Classic_and_Cult_TV.values)
plt.show()
```



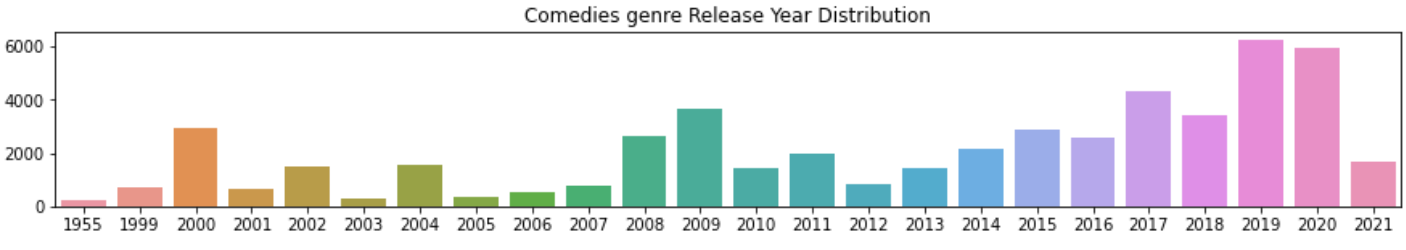
In [24]:
```
Classic_Movies = df['release_year'][' Classic Movies']
```

In [25]:
```
plt.title('Classic_Movies genre Release Year Distribution')
sns.barplot(x=Classic_Movies.index.values,y=Classic_Movies.values)
plt.show()
```



In [26]:
```
Comedies = df['release_year'][' Comedies']
```

In [27]:
```
plt.figure(figsize=(15,2))
plt.title('Comedies genre Release Year Distribution')
sns.barplot(x=Comedies.index.values,y=Comedies.values)
plt.show()
```



In [ ]:

In [28]:
```
data.head()
```

Out[28]:

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	cast
0	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Arno Greeff
1	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Chi Mhende
2	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Cindy Mahlangu
3	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Dillon Windvogel
4	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Duane Williams

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

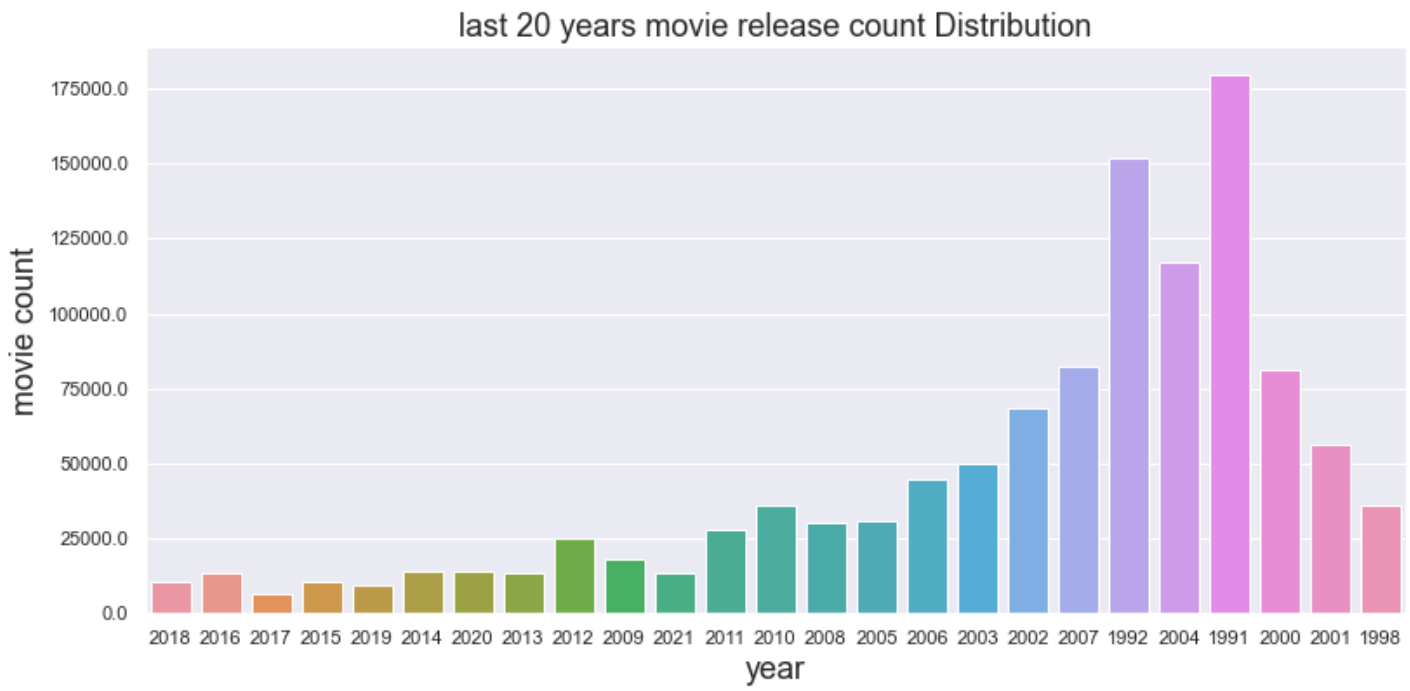
In [29]:
```
movies = data.loc[data['type']=='Movie']
```

In [30]:
```
last_20_year_movies_cnt = movies['release_year'].value_counts()[:25]
```

In [31]:
```
last_30_year_movies_cnt = movies['release_year'].value_counts()[:47]
```

In [32]:
```
sns.set(rc={"figure.figsize":(13, 6)})
sns.set(font_scale=1.5)
lst_20 = sns.barplot(x=last_20_year_movies_cnt.index.values , y = last_20_year_movies_cnt.
lst_20.set(xlabel='year',ylabel = "movie count", title ='last 20 years movie release count
lst_20.set_yticklabels(lst_20.get_yticks(), size =11)
lst_20.set_xticklabels(last_20_year_movies_cnt.index.values,size=11)
plt.show()
```

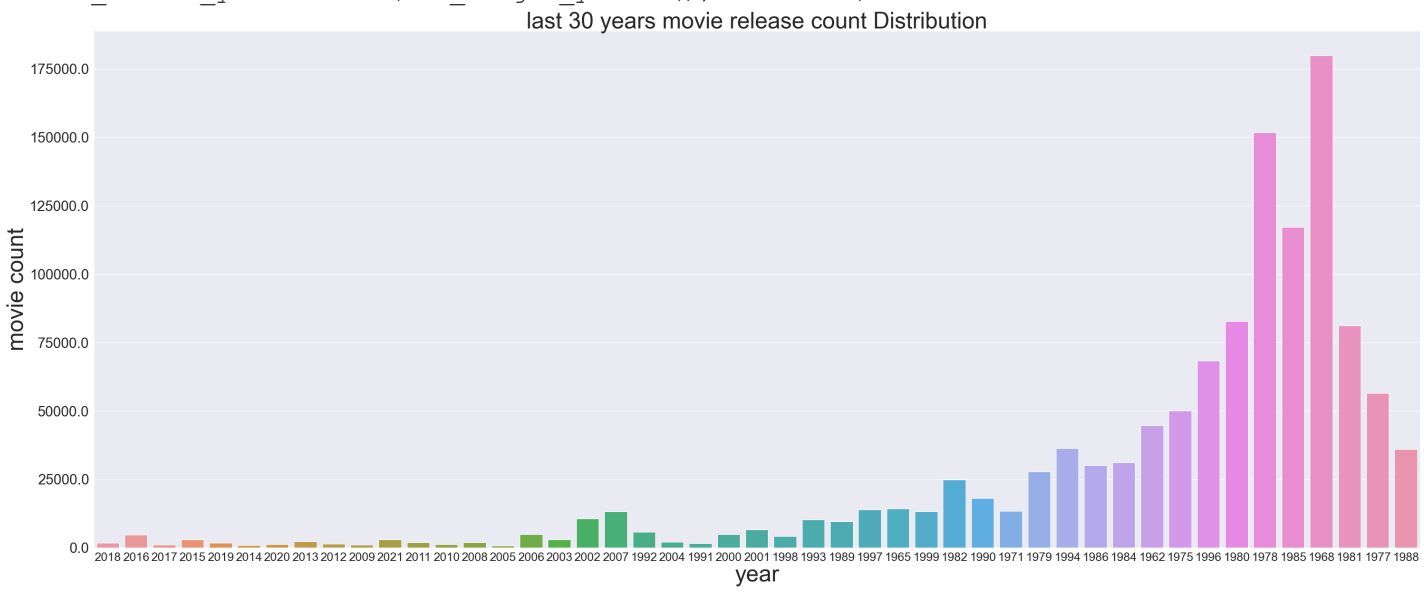
```
C:\Users\PC-DELL\AppData\Local\Temp\ipykernel_7148\4036817911.py:5: UserWarning: FixedForm
atter should only be used together with FixedLocator
  lst_20.set_yticklabels(lst_20.get_yticks(), size =11)
```



In [33]:

```
sns.set(rc={"figure.figsize":(50, 20)})
sns.set(font_scale=4)
lst_30 = sns.barplot(x=last_30_year_movies_cnt.index.values , y = last_30_year_movies_cnt.
lst_30.set(xlabel='year',ylabel = "movie count", title ='last 30 years movie release count
lst_30.set_yticklabels(lst_30.get_yticks(), size =30)
lst_30.set_xticklabels(last_30_year_movies_cnt.index.values,size=25)
plt.show()
```

```
C:\Users\PC-DELL\AppData\Local\Temp\ipykernel_7148\1634880165.py:5: UserWarning: FixedForm
atter should only be used together with FixedLocator
  lst_30.set_yticklabels(lst_30.get_yticks(), size =30)
```



In [ ]:

In [34]:

```
data
```

Out[34]:

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	
0	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	
1	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	
2	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	M
3	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	W
4	s2	TV Show	Blood & Water	2021	September 24, 2021	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	
...	...	...	...	...	...	...	...	...	...	...	...
1985046	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous	
1985047	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous	
1985048	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous	M
1985049	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous	R
1985050	s8798	TV Show	Zak Storm	2016	September 13, 2018	TV-Y7	3 Seasons	Kids' TV	Indonesia	anounimous	

1985051 rows × 11 columns

Tv Show Vs Movie Release date and Time Analysis

which year has maximum TV show released

In [108…

```
# This Analysis is about how Tv show released has changed over year
# It is observed that Tv show release has shown a gradual increase from 2015

# This will help business to understand why Tv show can be there main source of profit in
# As we can see the count is keep increasing, this shows that tv show will play a major po
```

In [35]:

```
tv_show_released= data.loc[data['type']=='TV Show']
```

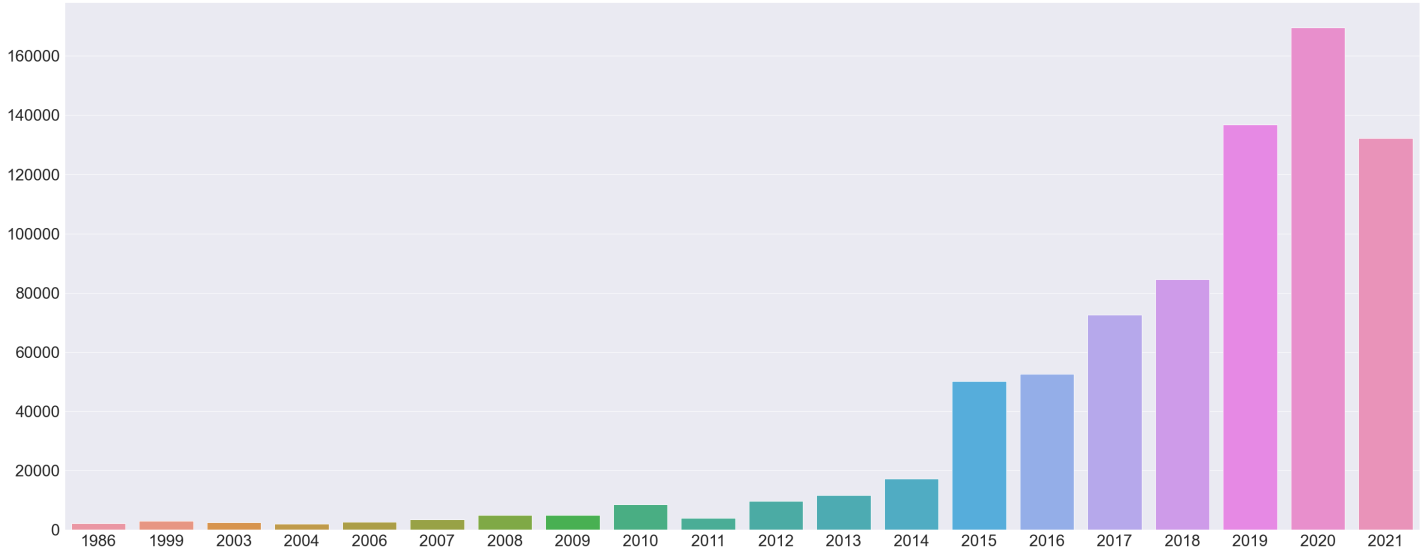
In [36]:

```
tv_show_released_count = tv_show_released['release_year'].value_counts()[:20]
tv_show_indx = tv_show_released_count.index.values
tv_show_values = tv_show_released_count.values
```

In [105…

```
sns.set(font_scale=3)
```

```
sns.barplot(x=tv_show_indx,y=tv_show_values,dodge=True)
plt.show()
```



Which month is good for tv show release

In [110…

```
# This Analysis is about which month can be beneficial for tv show stream for business.
# It is observed that august month most beneficial for tv show as compare to december mont
# This indicates that if most of tv shows which are released in august month more likely t
# as compare to other months.
```

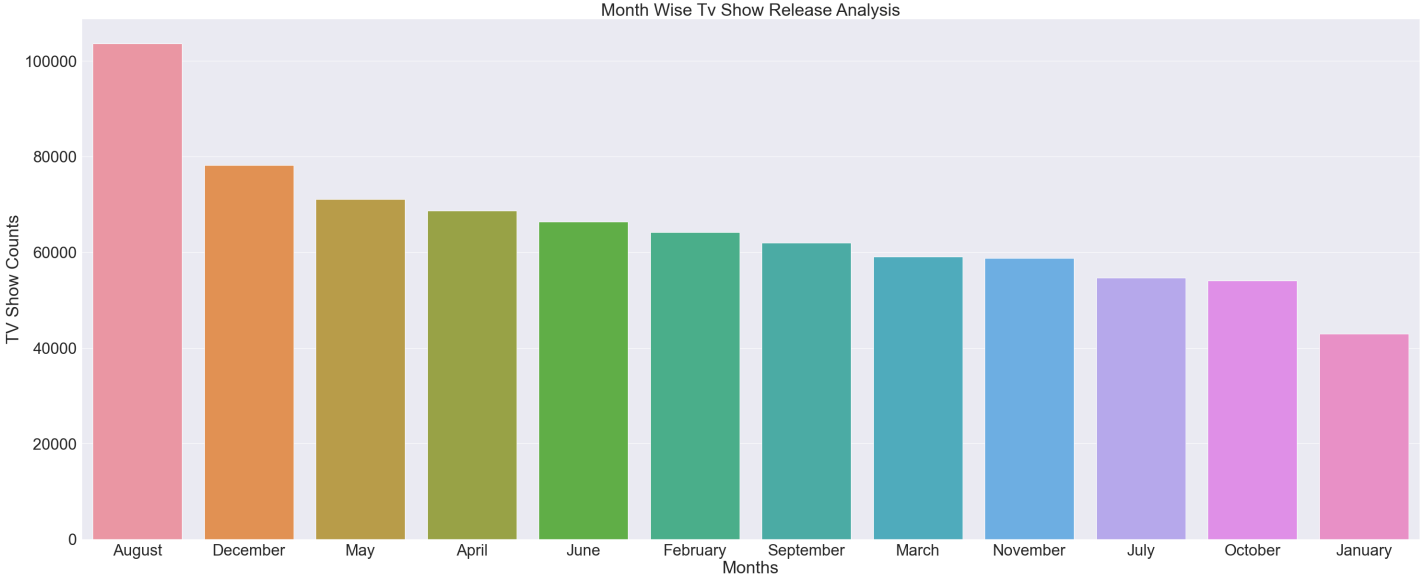
In [38]:

```
tv_show_released['date_added'] = tv_show_released['date_added'].astype('datetime64')
tv_show_month_names = tv_show_released['date_added'].dt.month_name()
most_occuring_tv_show_month = tv_show_month_names.value_counts()

most_occuring_tv_show_month_indx = most_occuring_tv_show_month.index.values
most_occuring_tv_show_month_values = most_occuring_tv_show_month.values
```

```
C:\Users\PC-DELL\AppData\Local\Temp\ipykernel_7148\3773642404.py:1: SettingWithCopyWarnin
g:
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_gu
ide/indexing.html#returning-a-view-versus-a-copy
  tv_show_released['date_added'] = tv_show_released['date_added'].astype('datetime64')
```

In [39]:

```
sns.set(font_scale=3)
month_wise_tv_show_release = sns.barplot(x=most_occuring_tv_show_month_indx,y=most_occurin
month_wise_tv_show_release.set(xlabel='Months',ylabel='TV Show Counts',title='Month Wise T
plt.show()
```



Comparison between tv show release on weekdays vs weekends

In [111…

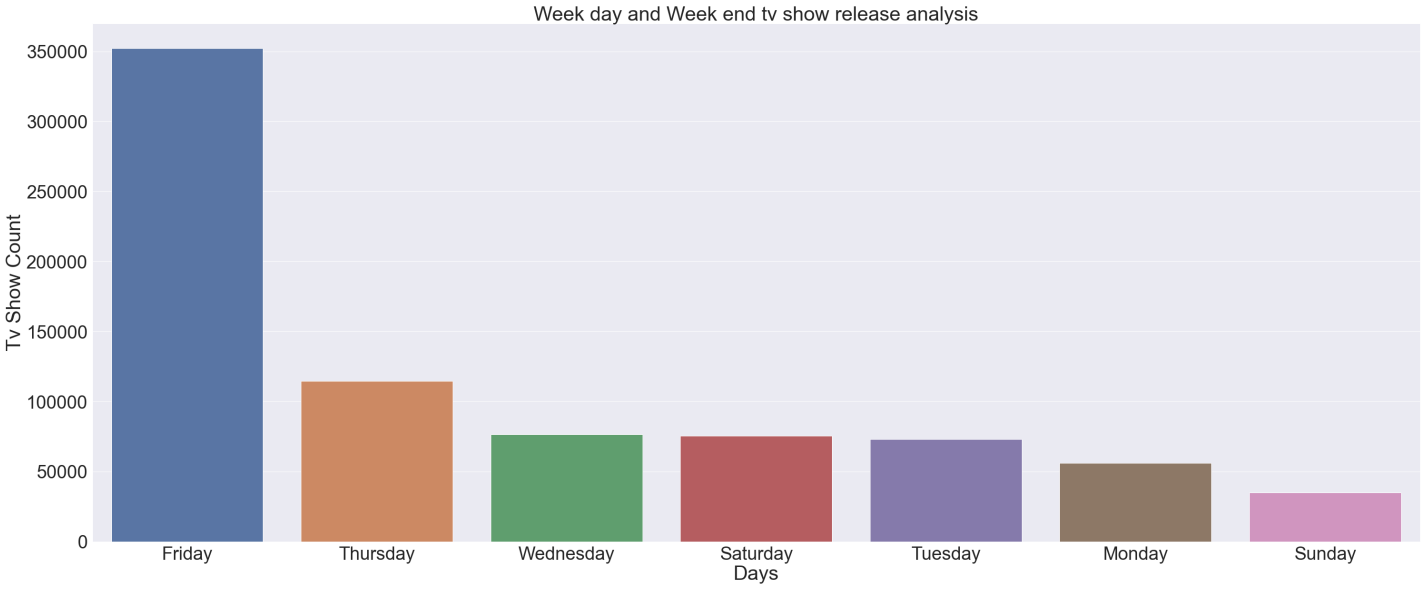
```
# This Analysis is about count differences between tv show release on weekdays vs weekends
# as per analysis it is seen that most of the tv shows are released on friday as compare t
# In future this values may change so in order make more profit in future it is advisable
# that they can stream tv show on saturday or sunday because those are days are mostly fre
```

In [40]:

```
tv_show_released_days = tv_show_released['date_added'].dt.day_name()
tv_show_released_days_counts = tv_show_released_days.value_counts()
tv_show_days_indx = tv_show_released_days_counts.index.values
tv_show_days_values = tv_show_released_days_counts.values
```

In [41]:

```
sns.set(font_scale=3.5)
day_wise_tv_show_release = sns.barplot(x=tv_show_days_indx,y=tv_show_days_values)
day_wise_tv_show_release.set(xlabel='Days',ylabel = "Tv Show Count", title ='Week day and
plt.show()
```



In [42]:

```
Movie_release = data.loc[data['type']=='Movie']
```

In [43]:
```
Movie_release
```

Out[43]:

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director
68042	s8	Movie	Sankofa	1993	September 24, 2021	TV-MA	125 min	Dramas	Ghana	Haile Gerima
68043	s8	Movie	Sankofa	1993	September 24, 2021	TV-MA	125 min	Dramas	Ghana	Haile Gerima
68044	s8	Movie	Sankofa	1993	September 24, 2021	TV-MA	125 min	Dramas	Ghana	Haile Gerima
68045	s8	Movie	Sankofa	1993	September 24, 2021	TV-MA	125 min	Dramas	Ghana	Haile Gerima
68046	s8	Movie	Sankofa	1993	September 24, 2021	TV-MA	125 min	Dramas	Ghana	Haile Gerima
...	...	...	...	...	...	...	...	...	...	...
1984997	s8789	Movie	You Carry Me	2015	July 1, 2016	TV-MA	157 min	International Movies	Montenegro	Ivona Juka
1984998	s8789	Movie	You Carry Me	2015	July 1, 2016	TV-MA	157 min	International Movies	Montenegro	Ivona Juka
1984999	s8789	Movie	You Carry Me	2015	July 1, 2016	TV-MA	157 min	International Movies	Montenegro	Ivona Juka
1985000	s8789	Movie	You Carry Me	2015	July 1, 2016	TV-MA	157 min	International Movies	Montenegro	Ivona Juka
1985001	s8789	Movie	You Carry Me	2015	July 1, 2016	TV-MA	157 min	International Movies	Montenegro	Ivona Juka

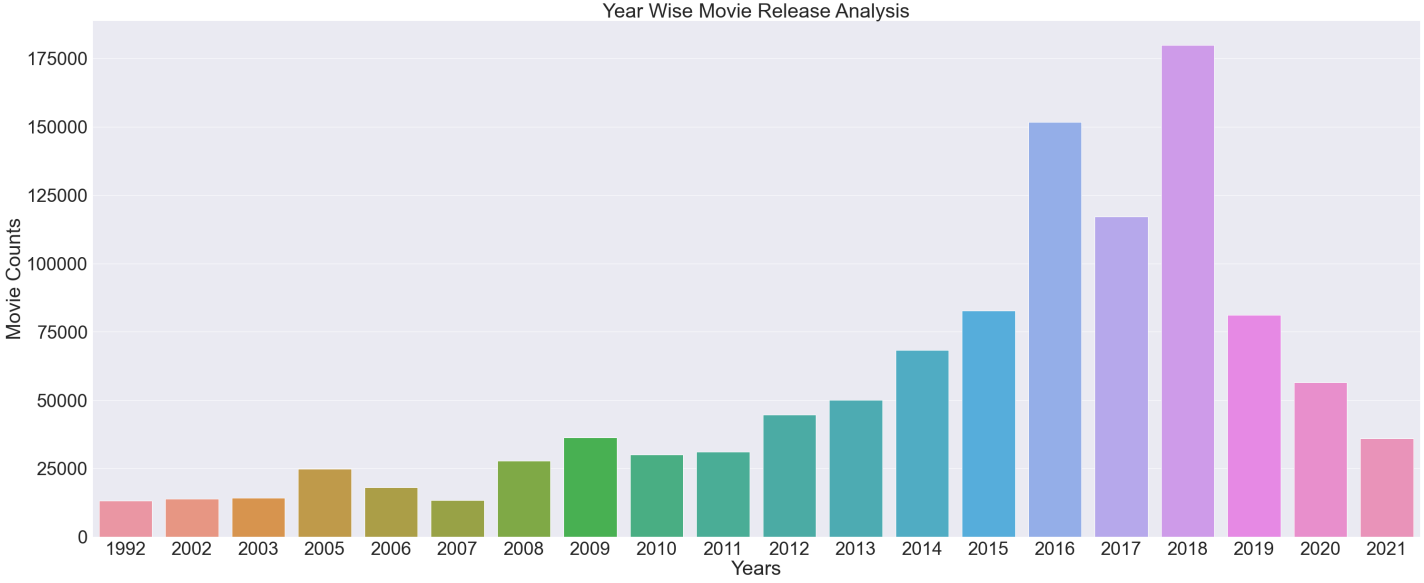
1201251 rows × 11 columns

which year has maximum Movie released

In [112…
```
# This Analysis is about how movie released has changed from 1992.
# It is observed that movie released has sharp increase in from 2015 and this peak is mai
# till 2018. As per analysis, movie released is has sharp decrease from 2018 to 2021
# This indicates that people are showing more on tv shows as compare to movie.
# It is advisable to to show case more movie on platform it will help business to get mor
```

In [44]:
```
movie_release_counts = Movie_release['release_year'].value_counts()[:20]
movie_release_counts_indx = movie_release_counts.index.values
movie_release_counts_values = movie_release_counts.values
```

In [45]:
```
year_wise_movie_counts = sns.barplot(x=movie_release_counts_indx,y=movie_release_counts_va
year_wise_movie_counts.set(xlabel='Years',ylabel='Movie Counts',title='Year Wise Movie Rel
plt.show()
```



Which month is good for Movie release

In [114…

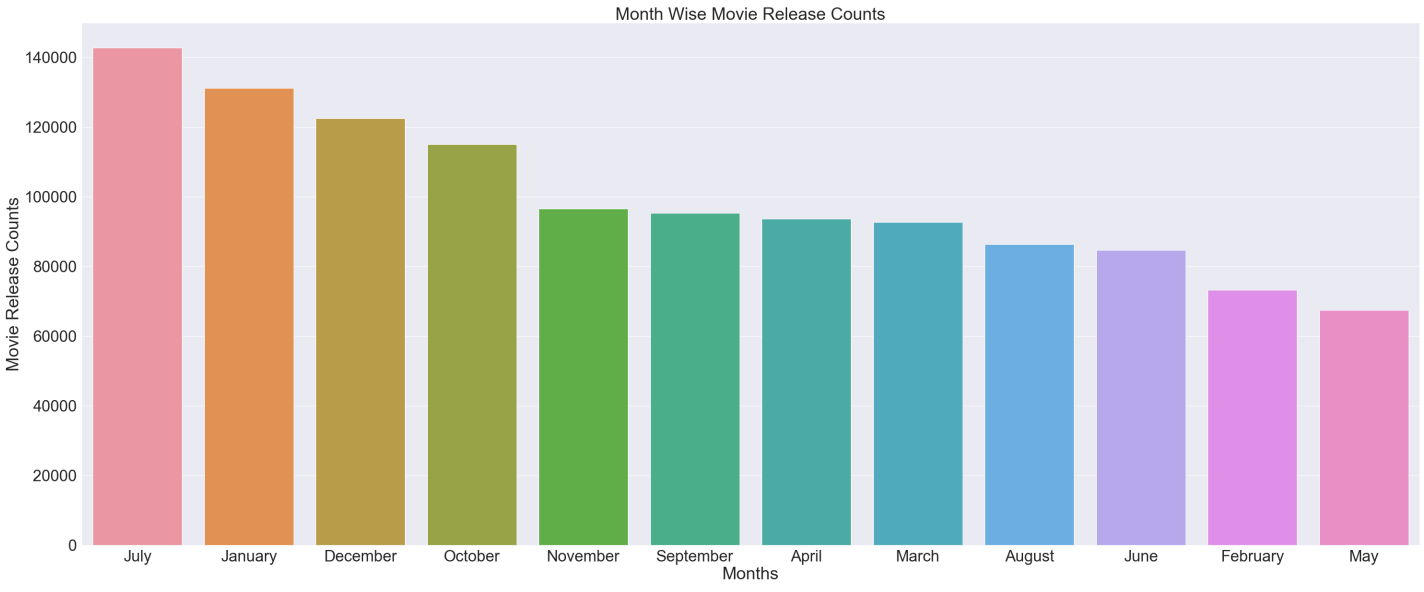
```
# This analysis is about which month is more suitable for movie release
# It is observed that july,december,october,january can be more suitable month for movie r
# as per analysis it stats that subscribers are likely to watch movies in these months, Th
```

In [46]:

```
Movie_release_years = Movie_release['date_added'].astype('datetime64')
Movie_release_months = Movie_release_years.dt.month_name()
Movie_release_months_release_counts = Movie_release_months.value_counts()
Movie_release_months_release_indx = Movie_release_months_release_counts.index.values
Movie_release_months_release_values = Movie_release_months_release_counts.values
```

In [47]:

```
sns.set(font_scale=3)
month_wise_movie_release = sns.barplot(x = Movie_release_months_release_indx , y = Movie_r
month_wise_movie_release.set(xlabel='Months', ylabel='Movie Release Counts' , title = 'Mon
plt.show()
```



In [ ]:

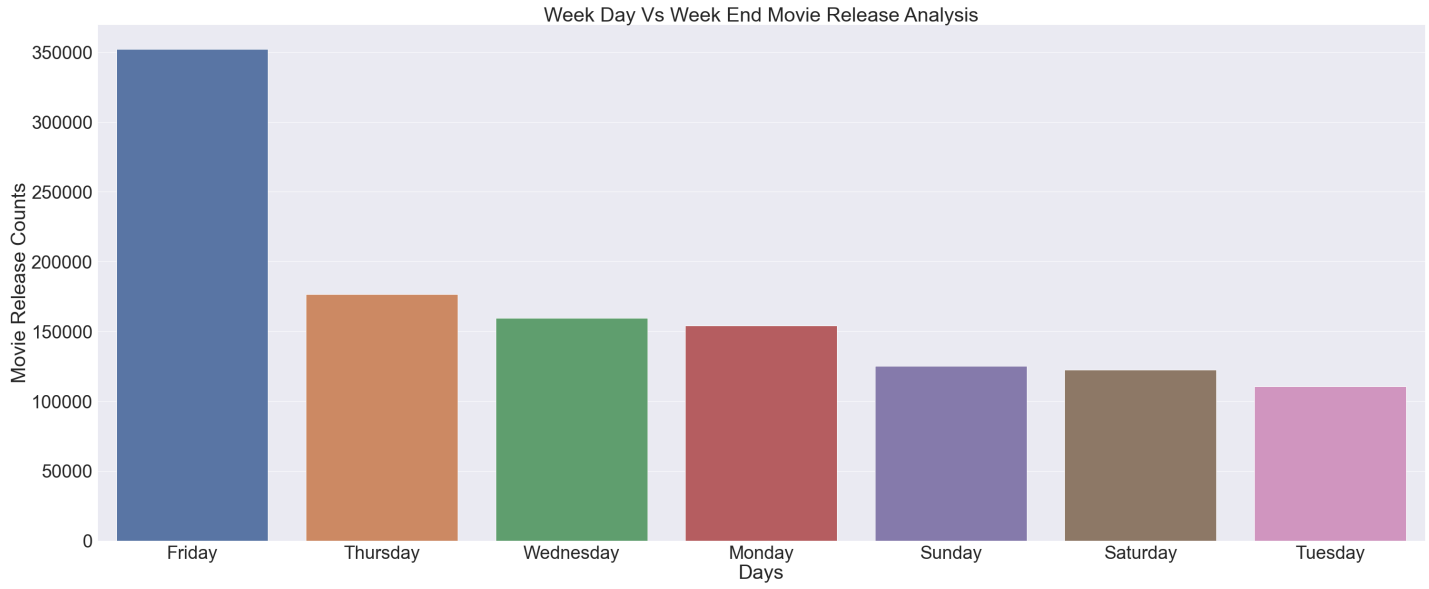
Comparison between Movie release on weekdays vs weekends

In [116…]
```
# This Analysis is about how movie release counts are different on weekdays vs weekends
# as per analysis it is observed that movie is most likely to be released on friday as cor
# Is is advisable to business to put more movies on weekends this will help business get r
```

In [48]:
```
Movie_release_weeks = Movie_release_years.dt.day_name()
```

In [49]:
```
Movie_release_weeks_counts = Movie_release_weeks.value_counts()
Movie_release_weeks_count_indx = Movie_release_weeks_counts.index.values
Movie_release_weeks_count_values = Movie_release_weeks_counts.values
```

In [115…]
```
sns.set(font_scale=3.5)
day_wise_tv_show_release = sns.barplot(x=Movie_release_weeks_count_indx,y=Movie_release_we
day_wise_tv_show_release.set(xlabel='Days',ylabel = "Movie Release Counts", title ='Week D
plt.show()
```



In [ ]:

In [ ]:

In [51]:
```
data['date_added'] = pd.to_datetime(data['date_added'])
data['date_added_year'] = data['date_added'].dt.year
data['date_added_week'] = data['date_added'].dt.day_name()
data['date_added_day'] = data['date_added'].dt.day
data['date_added_month'] = data['date_added'].dt.month
```

In [52]:
```
data.head()
```

Out[52]:

show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	cast

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	cast
0	s2	TV Show	Blood & Water	2021	2021-09-24	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Arno Greeff
1	s2	TV Show	Blood & Water	2021	2021-09-24	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Chi Mhende
2	s2	TV Show	Blood & Water	2021	2021-09-24	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Cindy Mahlangu
3	s2	TV Show	Blood & Water	2021	2021-09-24	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Dillon Windvogel
4	s2	TV Show	Blood & Water	2021	2021-09-24	TV-MA	2 Seasons	International TV Shows	South Africa	anounimous	Duane Williams

India Vs USA Data Analysis

In [54]:
```
india_data = data.loc[data['country']=='India']
```

In [55]:
```
india_data
```

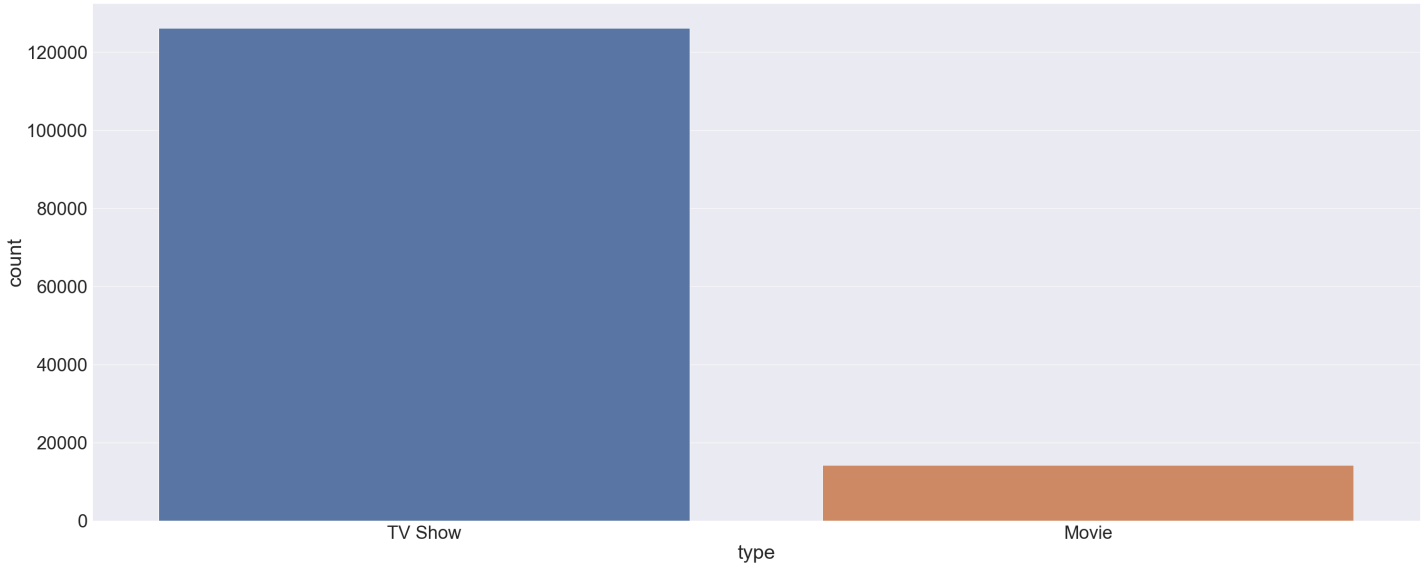
Out[55]:

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	
12372	s5	TV Show	Kota Factory	2021	2021-09-24	TV-MA	2 Seasons	International TV Shows	India	B. V. Nandini Reddy	
12373	s5	TV Show	Kota Factory	2021	2021-09-24	TV-MA	2 Seasons	International TV Shows	India	B. V. Nandini Reddy	K
12374	s5	TV Show	Kota Factory	2021	2021-09-24	TV-MA	2 Seasons	International TV Shows	India	B. V. Nandini Reddy	
12375	s5	TV Show	Kota Factory	2021	2021-09-24	TV-MA	2 Seasons	International TV Shows	India	B. V. Nandini Reddy	
12376	s5	TV Show	Kota Factory	2021	2021-09-24	TV-MA	2 Seasons	International TV Shows	India	B. V. Nandini Reddy	Kak
...	...	...	...	...	...	...	...	...	...	...	...
1895504	s6578	Movie	Deadline: Sirf 24 Ghante	2006	2019-12-31	TV-14	97 min	Independent Movies	India	Tanveer Khan	
1895505	s6578	Movie	Deadline: Sirf 24 Ghante	2006	2019-12-31	TV-14	97 min	Independent Movies	India	Tanveer Khan	
1895506	s6578	Movie	Deadline: Sirf 24 Ghante	2006	2019-12-31	TV-14	97 min	Independent Movies	India	Tanveer Khan	S

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	
1895507	s6578	Movie	Deadline: Sirf 24 Ghante	2006	2019-12-31	TV-14	97 min	Independent Movies	India	Tanveer Khan	
1895508	s6578	Movie	Deadline: Sirf 24 Ghante	2006	2019-12-31	TV-14	97 min	Independent Movies	India	Tanveer Khan	

140406 rows × 15 columns

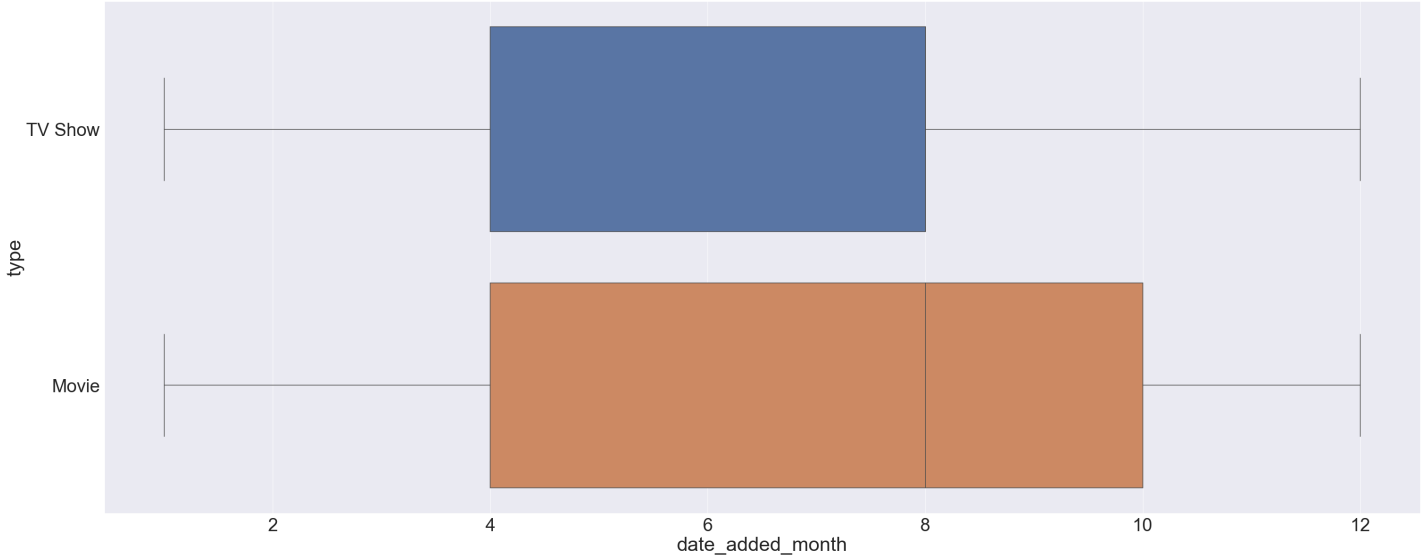
In [57]:
```
# In India public is more tend to watch tv show
sns.countplot(x = india_data['type'])
plt.show()
```



Month Wise TV show vs Movie Release on platform Analysis

In [58]:
```
# Movie :  Mostly Released from april to october get released
# TV show : Mostly Released from april to august

sns.boxplot(x = india_data['date_added_month'],y=india_data['type'])
plt.plot()
```

Out[58]:
```
[]
```

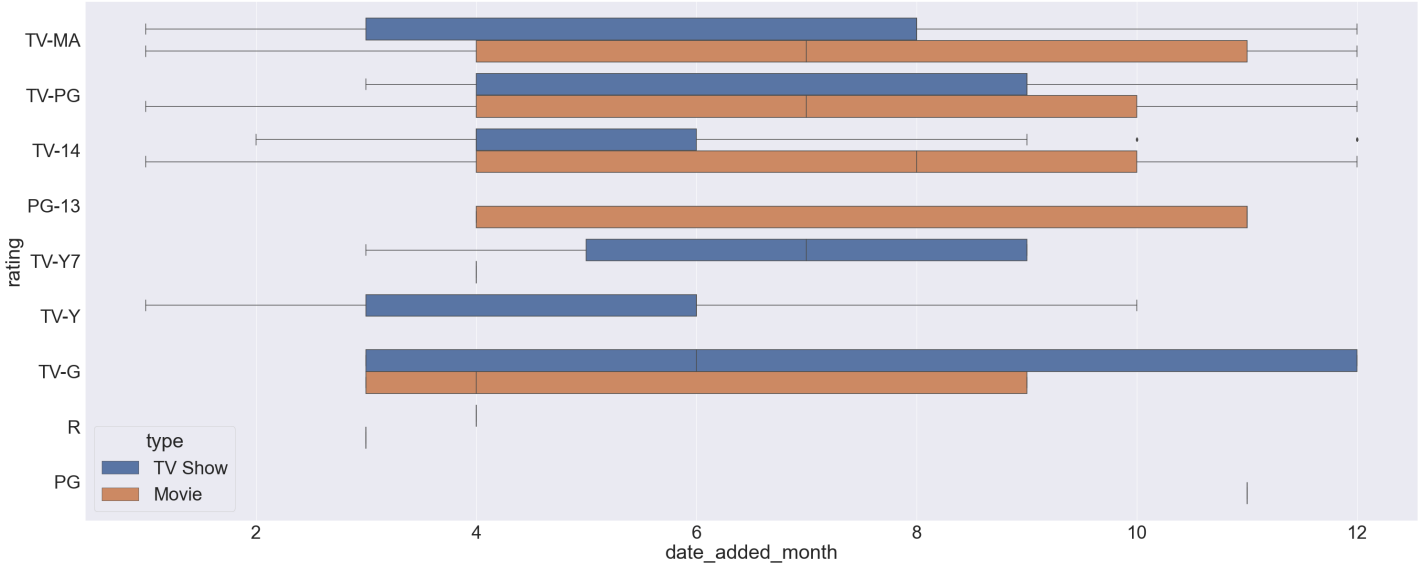


In [59]:

```
# Rating Meaning :
# TV-MA : This program is specifically designed to be viewed by adults and therefore may b
# TV-PG : parental guidance is recommended; these programs may be unsuitable for younger c
# TV-14 : these shows may be unsuitable for children under 14
# PG-13 :Parents strongly cautioned.
# TV-Y7 : programs most appropriate for children age 7 and up
# TV-Y : programs aimed at a very young audience, including children from ages 2-6
# TV-G :programs suitable for all ages
# R : Under 17 requires accompanying parent or adult guardian
# PG : PARENTAL GUIDANCE SUGGESTED. Some material may not be suitable for children

##########################################################################################

# Postive side Organisation is focusing more on content for all age and focusing less on b

# data depicts that TV-G rating tv shows are released more on India as compare to movies.
# Reason : Public is looking to content which is suitable for all and family can also watc
# target release could be around june month because of summer vacation people look for mo


# Tv -14 : data depicts that under 14 age watchers are more interested in watching movies
# Best for release could be near around month of august


sns.boxplot(x = india_data['date_added_month'],y=india_data['rating'],hue=india_data['type
plt.plot()
```

Out[59]:

```
[]
```

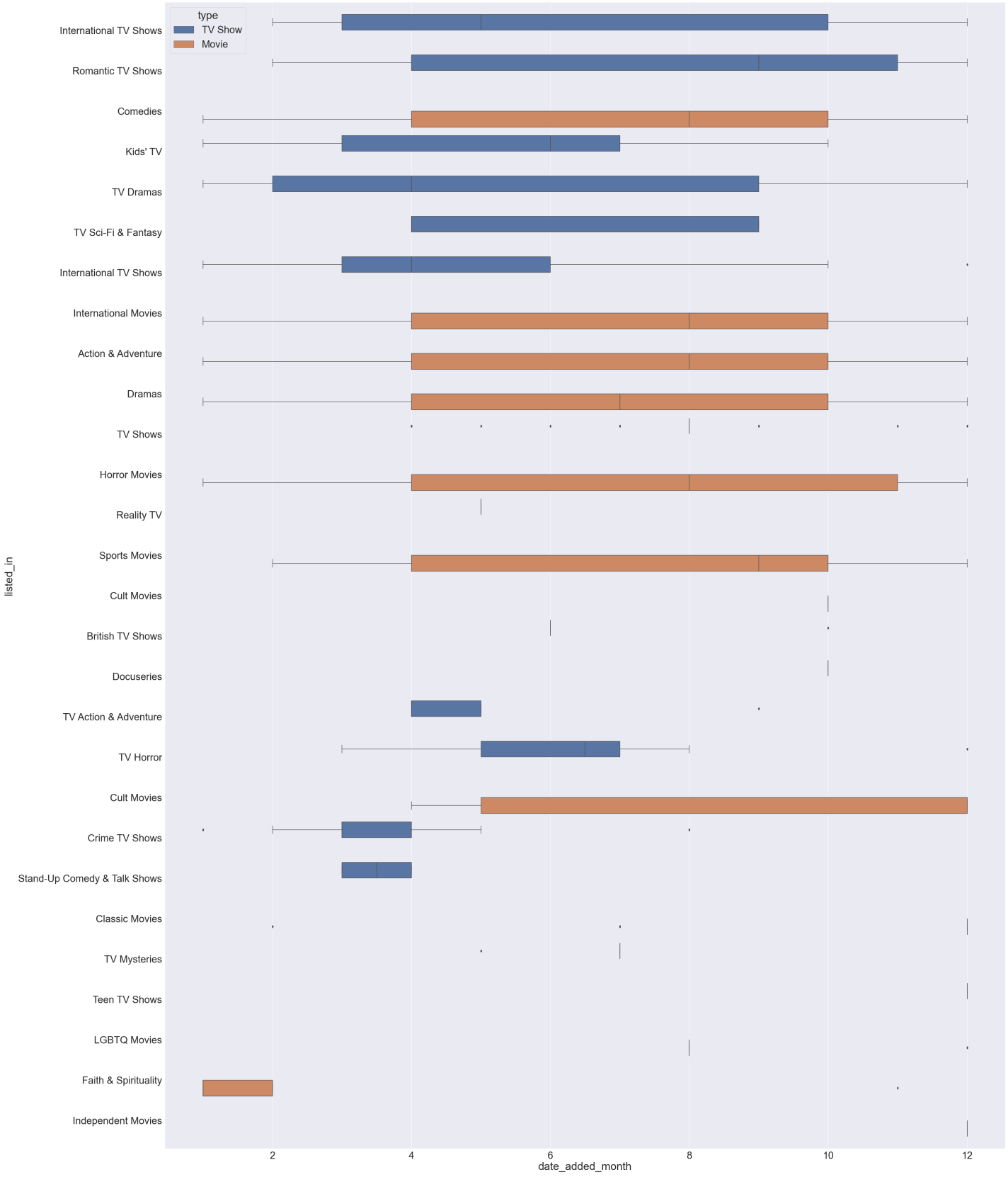


In [60]:

```
# data depicts that average release of romantic tv shows are more than international tv sh
# data depicts that average release of sports movies are more than Horror which is more pi
# watchers will get more enlightened with country's sport history

plt.figure(figsize=(50,70))
sns.set(font_scale=3)
genre_wise_date_added_analysis = sns.boxplot(x = india_data['date_added_month'],y=india_da
plt.plot()
```

Out[60]:

```
[]
```

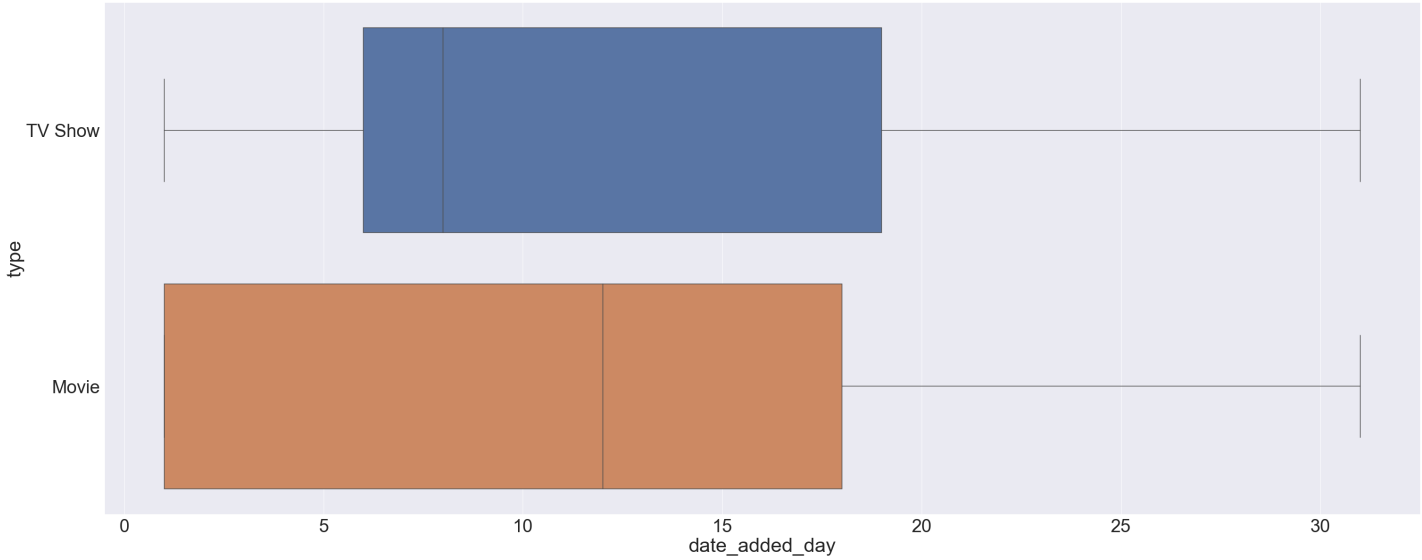


Day Wise Analysis for TV show Vs Movie Release on Platform

In [117…

```
# Data depicts that average release of movie is ranging around 10th day to 15th day of a r
# Where tv show average release is ranging from 5th day to 14th day of month.

# Organisation can target movie/tv show release around mentioned days for better audience
```

```
sns.boxplot(x = india_data['date_added_day'],y=india_data['type'])
plt.show()
```



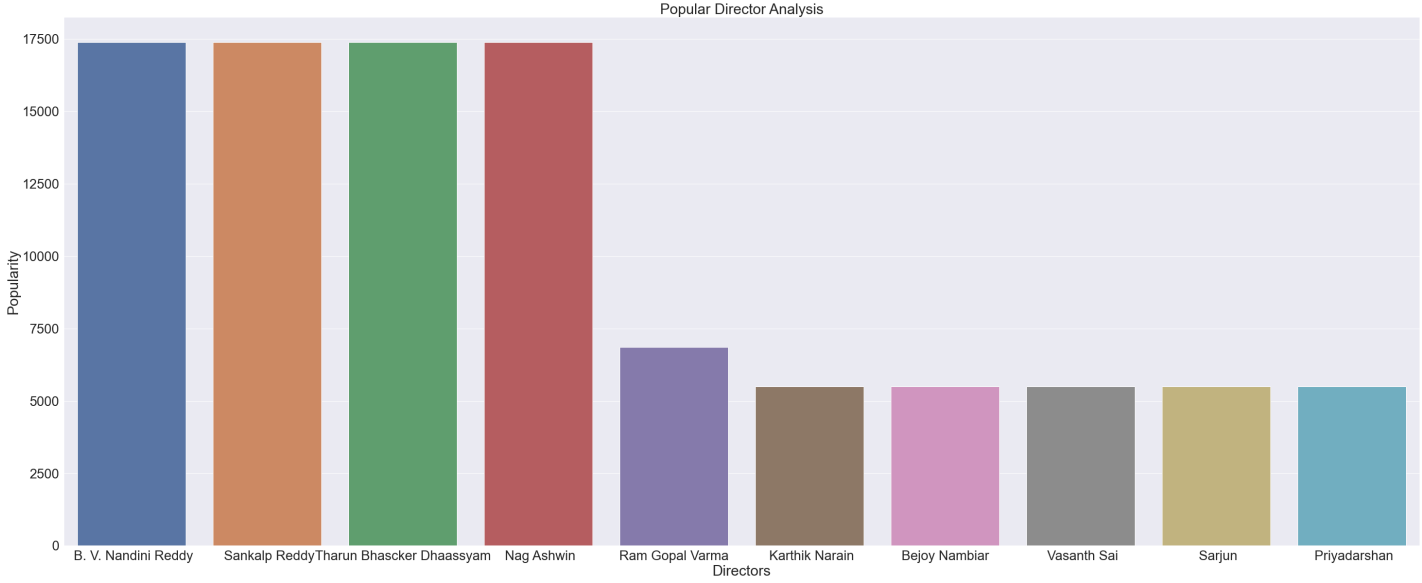
Popular Director in india

In [63]:
```
india_data['director']
```

Out[63]:
```
12372       B. V. Nandini Reddy
12373       B. V. Nandini Reddy
12374       B. V. Nandini Reddy
12375       B. V. Nandini Reddy
12376       B. V. Nandini Reddy
                   ...
1895504            Tanveer Khan
1895505            Tanveer Khan
1895506            Tanveer Khan
1895507            Tanveer Khan
1895508            Tanveer Khan
Name: director, Length: 140406, dtype: object
```

In [64]:
```
popular_director = india_data[india_data['director'].apply(lambda x : x!='anounimous')]['d
popular_director_indx = popular_director.index.values
popular_director_values = popular_director.values
```

In [118…
```
# This Analysis will help business to understand which director is impacting more audience
# it will give more profit if people will find those director movie/tv show directed by sp

plt.figure(figsize=(50,20))
sns.set(font_scale=2.5)
popular_director_analysis = sns.barplot(x=popular_director_indx,y=popular_director_values)
popular_director_analysis.set(xlabel='Directors',ylabel = "Popularity", title ='Popular Di
plt.show()
```

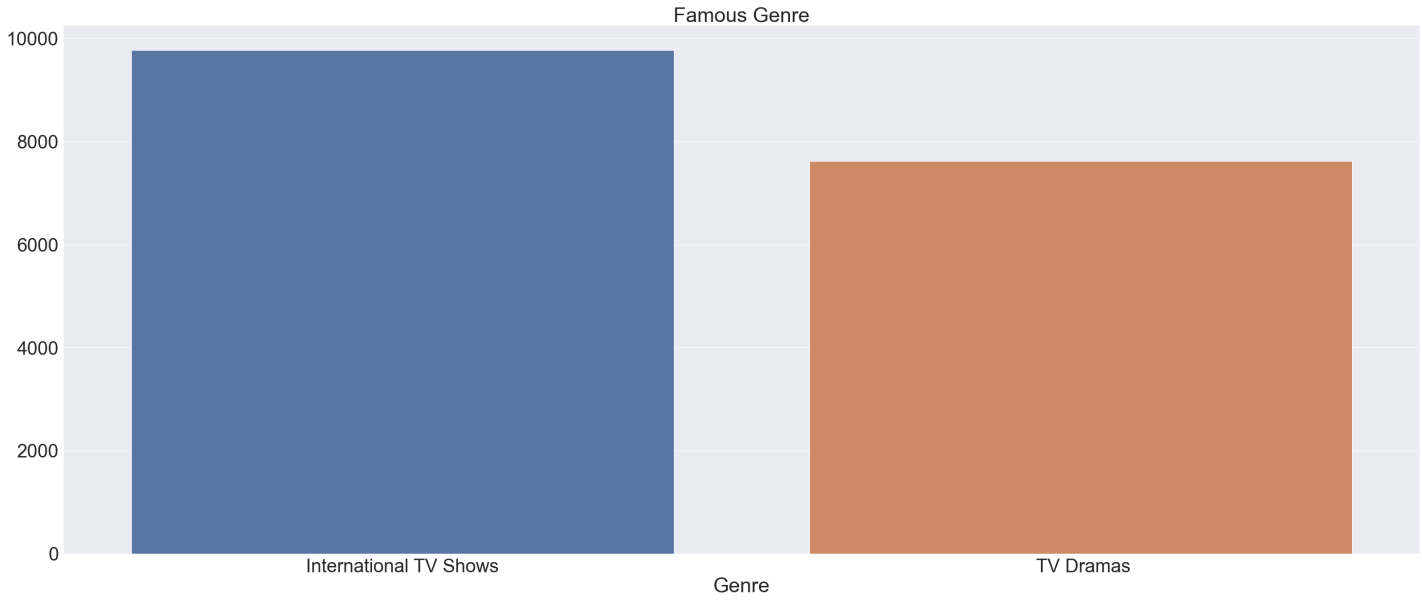


Which genre is mostly worked by popular director

In [66]:
```
india_popular_director_genre = india_data.loc[india_data['director']==' B. V. Nandini Redd
india_popular_director_genre_indx = india_popular_director_genre['listed_in'].value_counts
india_popular_director_genre_values = india_popular_director_genre['listed_in'].value_cour
```

In [121…
```
# This Analysis will be helpful for business to understand which genre to showcase more wh
# is about to stream on platform.
# This will help business to get more audience coverage.
```

In [119…
```
sns.set(font_scale=(3.5))
india_popular_director_genre_plot = sns.barplot(x=india_popular_director_genre_indx,y=indi
india_popular_director_genre_plot.set(xlabel='Genre',title='Famous Genre')
plt.show()
```



Most Popular Actors in India

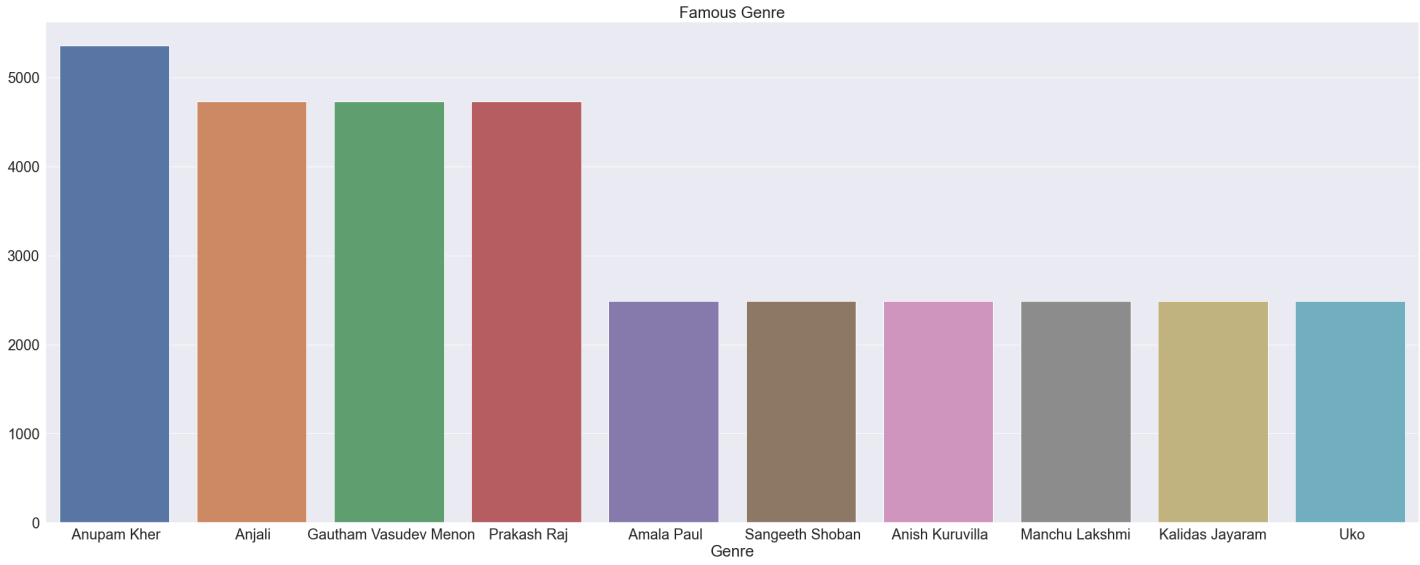
In [123…
```
# This will help business to understand how indian public is so specific towards actors be
# In which there favourite actors are present. This Analysis will help business to get mo
```

In [100…

```
Most_popular_actor_india = india_data['cast'].value_counts()[:10]
Most_popular_actor_india_indx = Most_popular_actor_india.index.values
Most_popular_actor_india_vals = Most_popular_actor_india.values
```

In [99]:

```
plt.figure(figsize=(40,15))
sns.set(font_scale=(2.15))
india_popular_director_genre_plot = sns.barplot(x=Most_popular_actor_india_indx,y=Most_pop
india_popular_director_genre_plot.set(xlabel='Genre',title='Famous Genre')
plt.show()
```



In [124…

```
us_data = data.loc[data['country']=='United States']
```

In [125…

```
us_data
```

Out[125…

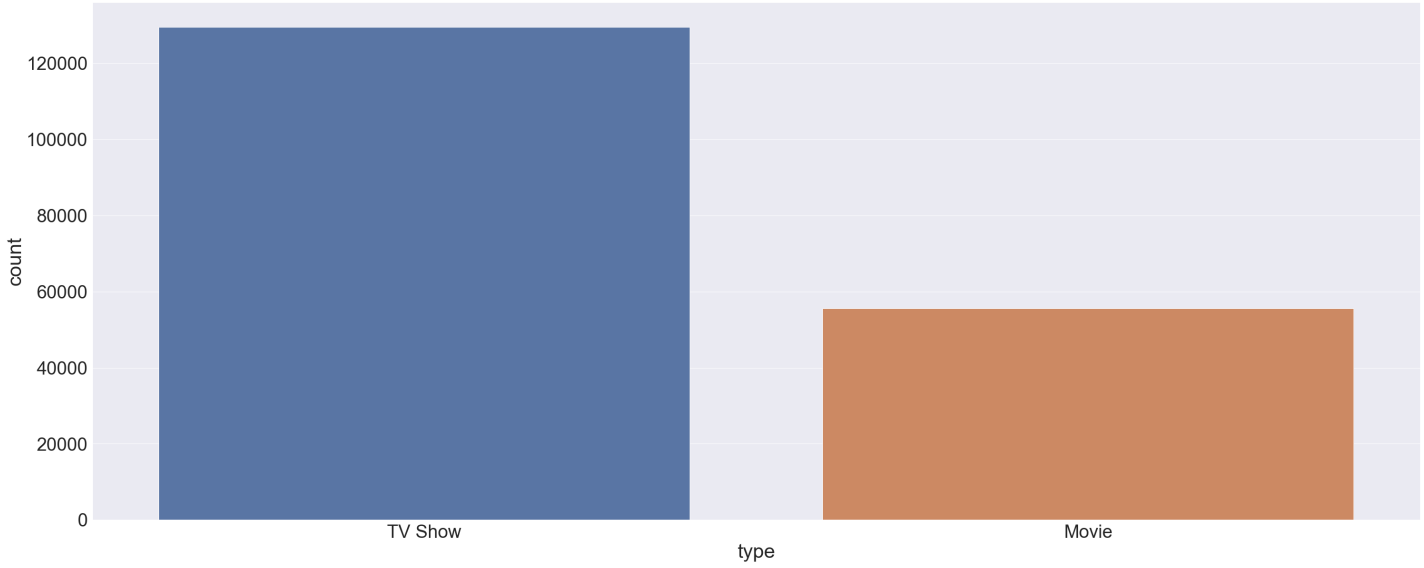
	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director	
2205	s3	TV Show	Ganglands	2021	2021-09-24	TV-MA	1 Season	Crime TV Shows	United States	Danny Cannon	
2206	s3	TV Show	Ganglands	2021	2021-09-24	TV-MA	1 Season	Crime TV Shows	United States	Danny Cannon	
2207	s3	TV Show	Ganglands	2021	2021-09-24	TV-MA	1 Season	Crime TV Shows	United States	Danny Cannon	
2208	s3	TV Show	Ganglands	2021	2021-09-24	TV-MA	1 Season	Crime TV Shows	United States	Danny Cannon	
2209	s3	TV Show	Ganglands	2021	2021-09-24	TV-MA	1 Season	Crime TV Shows	United States	Danny Cannon	
...	...	...	...	...	...	...	...	...	...	...	
1962845	s7742	Movie	Pierre Jackson	2018	2019-06-29	TV-MA	81 min	Romantic Movies	United States	Christopher Nolen	
1962846	s7742	Movie	Pierre Jackson	2018	2019-06-29	TV-MA	81 min	Romantic Movies	United States	Christopher Nolen	
1962847	s7742	Movie	Pierre Jackson	2018	2019-06-29	TV-MA	81 min	Romantic Movies	United States	Christopher Nolen	He

	show_id	type	title	release_year	date_added	rating	duration	listed_in	country	director
1962848	s7742	Movie	Pierre Jackson	2018	2019-06-29	TV-MA	81 min	Romantic Movies	United States	Christopher Nolen
1962849	s7742	Movie	Pierre Jackson	2018	2019-06-29	TV-MA	81 min	Romantic Movies	United States	Christopher Nolen

185137 rows × 15 columns

In [126…

```
# In US public is more tend to watch tv show
sns.countplot(x = us_data['type'])
plt.show()
```



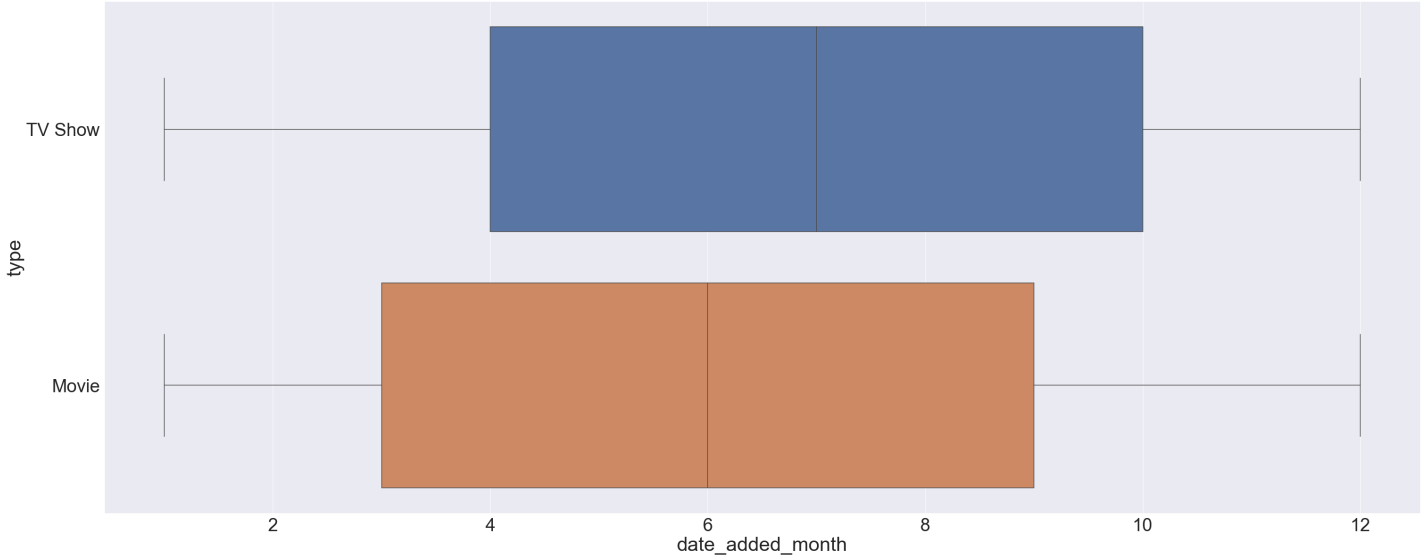
Month Wise TV show vs Movie Release on platform Analysis

In [129…

```
# It is observed that average tv show release is more than movie release on platform.
# This depicts that US audience is more likly to watch tv show as compare movie.
# This can main profit gainer where business can focus more.
sns.boxplot(x = us_data['date_added_month'],y=us_data['type'])
plt.plot()
```

Out[129…

```
[]
```



In [137…

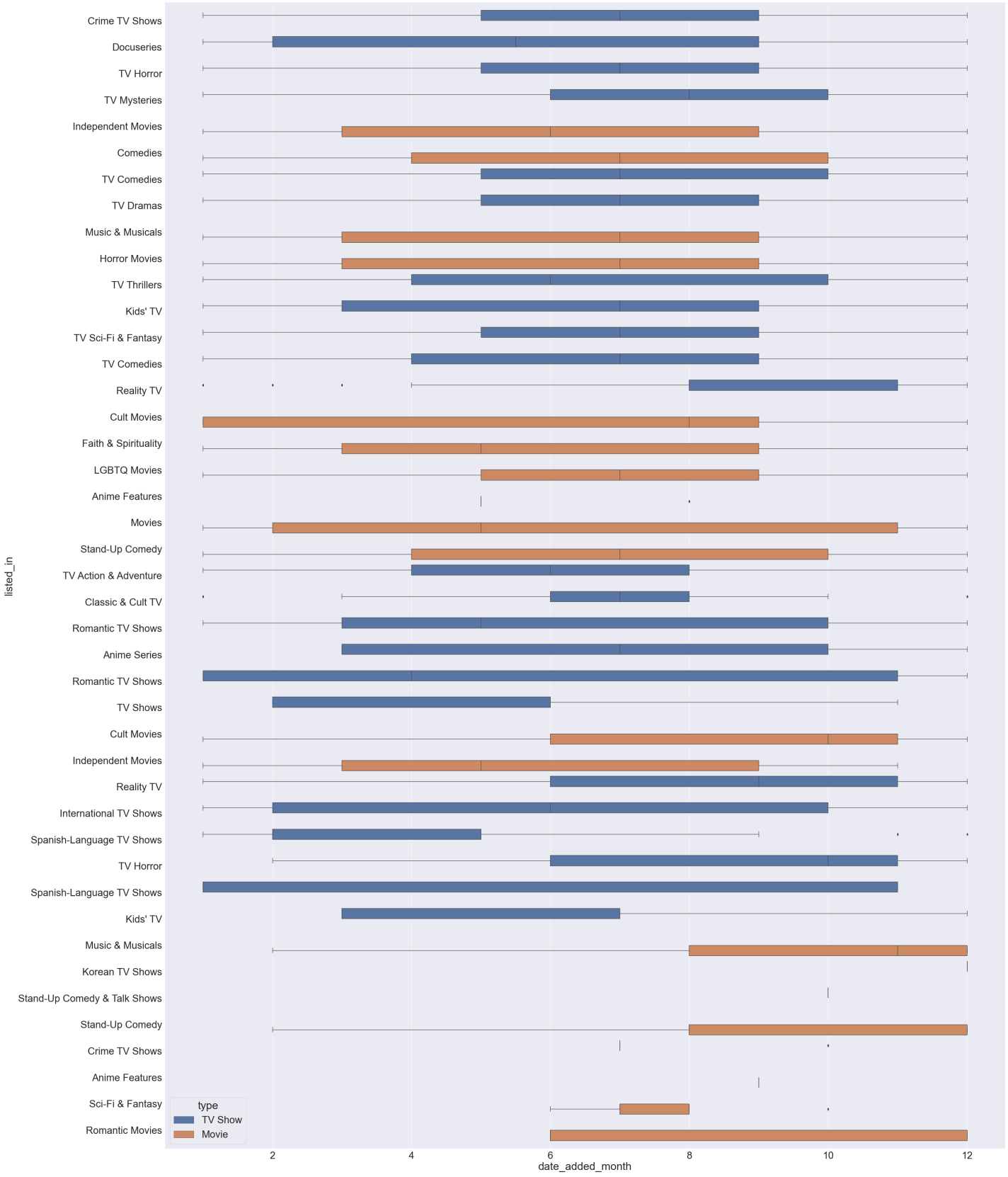
```
# data depicts that average release of romantic tv shows are less than international tv sh
# data depicts that US audience is not showing much interest in anime feature, crime tv sh
# of those content is less

# In order to cover more US audience focus more on other family and action based genre
```

In [136…

```
plt.figure(figsize=(50,70))
sns.set(font_scale=3)
genre_wise_date_added_analysis = sns.boxplot(x = us_data['date_added_month'],y=us_data['li
plt.plot()
```

Out[136…

```
[]
```



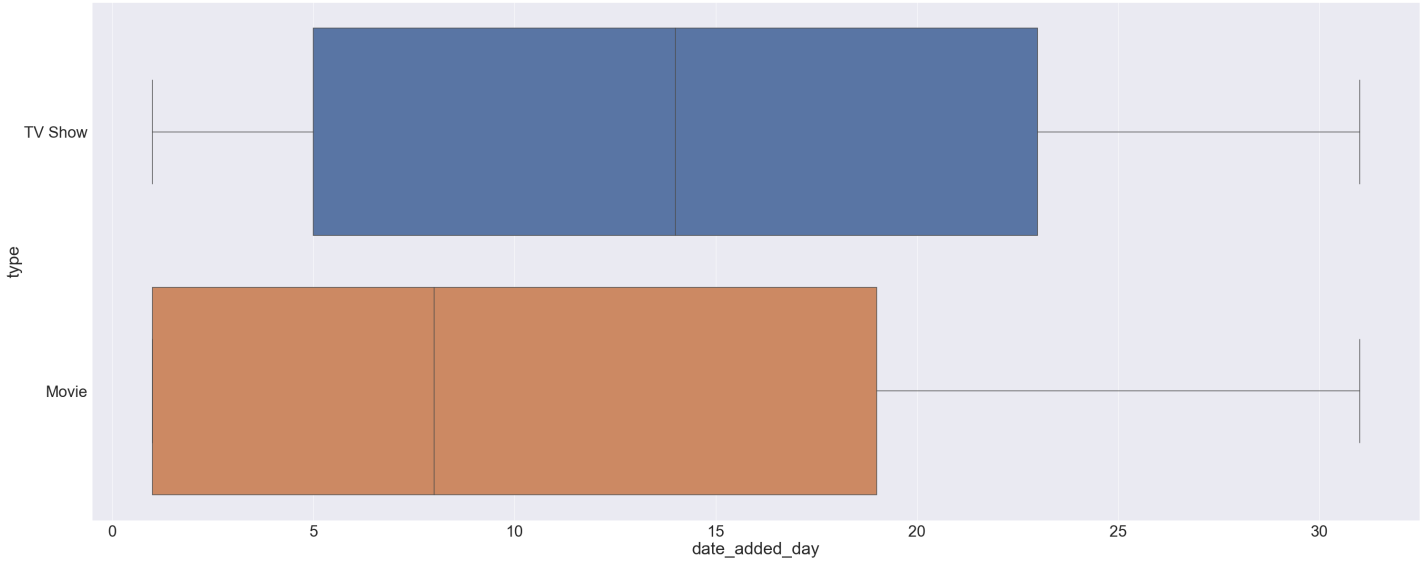
Day Wise Analysis for TV show Vs Movie Release on Platform

In [145…

```
# Data depicts that average release of movie is ranging around 5th day to 15th day of a mo
# Where tv show average release is ranging from 5th day to 14th day of month.

# Organisation can target movie/tv show release around mentioned days for better audience
```

In [140…

```
sns.boxplot(x = us_data['date_added_day'],y=us_data['type'])
plt.show()
```



Popular Director in US

In [143…

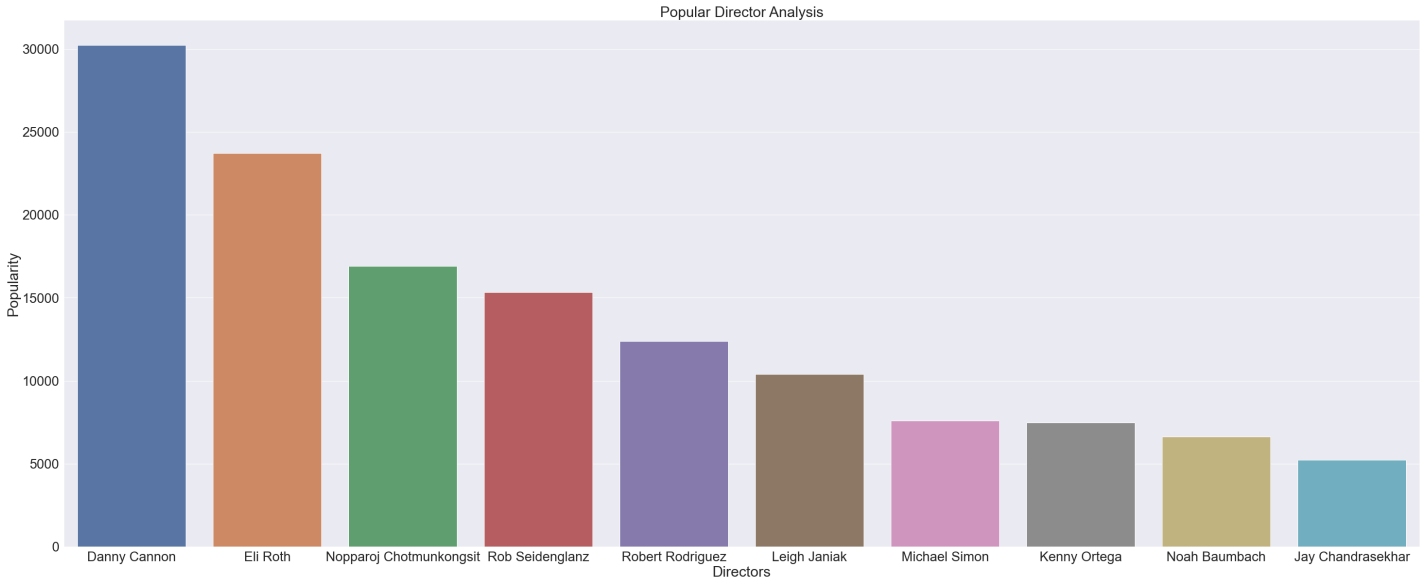
```
popular_director = us_data[us_data['director'].apply(lambda x : x!='anounimous')]['directo
popular_director_indx = popular_director.index.values
popular_director_values = popular_director.values
```

In [146…

```
# This Analysis will help business to understand which director is impacting more audience
# it will give more profit if people will find those director movie/tv show directed by sp
```

In [144…

```
plt.figure(figsize=(50,20))
sns.set(font_scale=2.5)
popular_director_analysis = sns.barplot(x=popular_director_indx,y=popular_director_values)
popular_director_analysis.set(xlabel='Directors',ylabel = "Popularity", title ='Popular Di
plt.show()
```

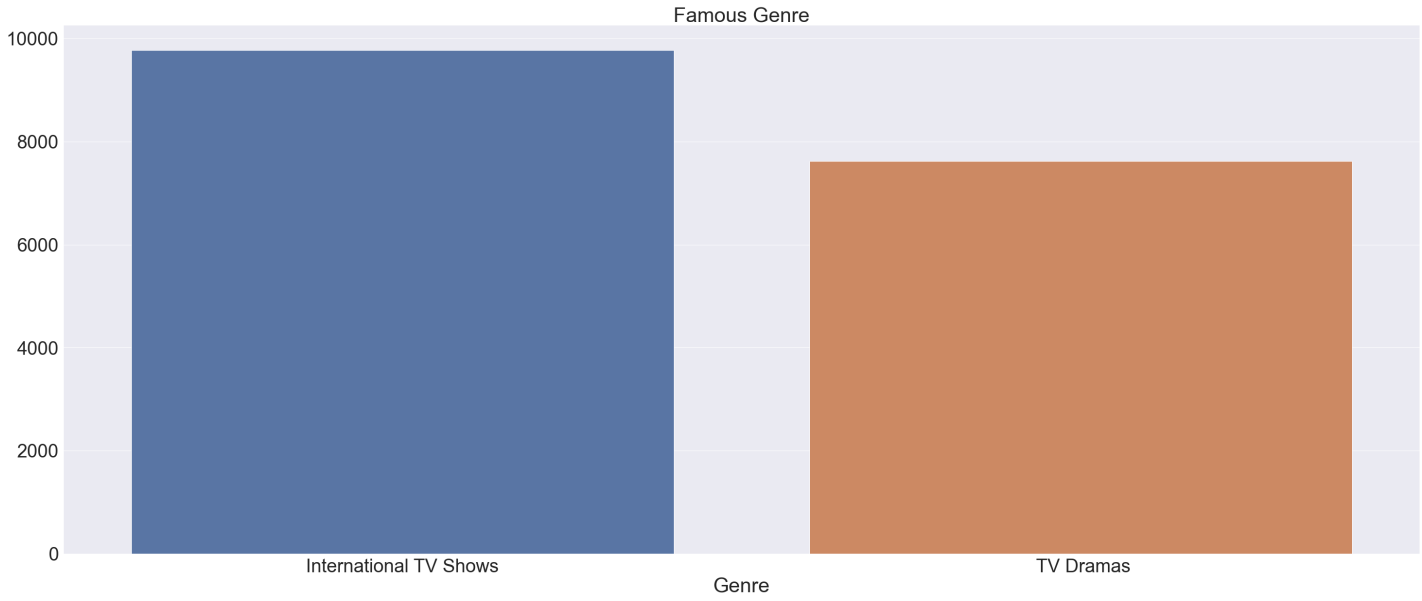


Which genre is mostly worked by popular director

In [154…
```
us_popular_director_genre = us_data.loc[us_data['director']=='Danny Cannon']
us_popular_director_genre_indx = us_popular_director_genre['listed_in'].value_counts().inc
us_popular_director_genre_values = us_popular_director_genre['listed_in'].value_counts().v
```

In [155…
```
# This Analysis will be helpful for business to understand which genre to showcase more wh
# is about to stream on platform.
# This will help business to get more audience coverage.
```

In [156…
```
sns.set(font_scale=(3.5))
india_popular_director_genre_plot = sns.barplot(x=india_popular_director_genre_indx,y=indi
india_popular_director_genre_plot.set(xlabel='Genre',title='Famous Genre')
plt.show()
```

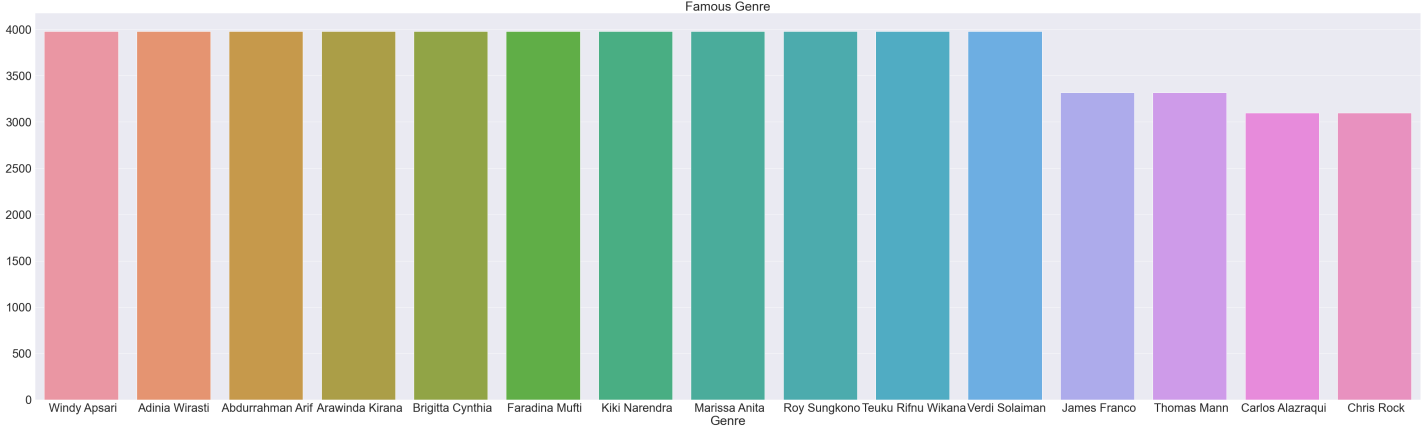


Most Popular Actors in US

In [157…
```
# This will help business to understand how indian public is so specific towards actors be
# In which there favourite actors are present. This Analysis will help business to get mo
```

In [178…
```
Most_popular_actor_us = us_data['cast'].value_counts()[:15]
Most_popular_actor_us_indx = Most_popular_actor_us.index.values
Most_popular_actor_us_vals = Most_popular_actor_us.values
```

In [188…
```
sns.set(rc={"figure.figsize":(70, 20)})
sns.set(font_scale=(3))
us_popular_director_genre_plot = sns.barplot(x=Most_popular_actor_us_indx,y=Most_popular_a
us_popular_director_genre_plot.set(xlabel='Genre',title='Famous Genre')
plt.show()
```



Recommendations

In [193…

```
# 1. Audience is more like to watch tv show , this is what we found in analysis
# 2. Audience is more looking for content which can be watched with family(specificly in i
# 3. Audience is more like to watch tv show/ movies which have there favourite director ar
# 4. It is observed that there is gradual decrease in movies in from 2019 which indicate a
# in near future this can be an alarming condition for business so better to maintain movi

# It is recommended that to put content on platform in between friday to sunday because it
# It is recommended that to put more tv show in month of august and movies in month of jul
```

In [199…

```
Collecting pyppeteer
  Downloading pyppeteer-1.0.2-py3-none-any.whl (83 kB)
     ---------------------------------------- 83.4/83.4 kB 1.2 MB/s eta 0:00:00
Requirement already satisfied: importlib-metadata>=1.4 in c:\users\pc-dell\anaconda3\envs
\tf_gpu\lib\site-packages (from pyppeteer) (4.8.2)
Collecting websockets<11.0,>=10.0
  Downloading websockets-10.4-cp310-cp310-win_amd64.whl (101 kB)
     ---------------------------------------- 101.4/101.4 kB 2.9 MB/s eta 0:00:00
Collecting certifi>=2021
  Downloading certifi-2022.9.24-py3-none-any.whl (161 kB)
     ---------------------------------------- 161.1/161.1 kB 1.4 MB/s eta 0:00:00
Requirement already satisfied: tqdm<5.0.0,>=4.42.1 in c:\users\pc-dell\anaconda3\envs\tf_g
pu\lib\site-packages (from pyppeteer) (4.63.1)
Requirement already satisfied: urllib3<2.0.0,>=1.25.8 in c:\users\pc-dell\anaconda3\envs\t
f_gpu\lib\site-packages (from pyppeteer) (1.26.9)
Collecting appdirs<2.0.0,>=1.4.3
  Downloading appdirs-1.4.4-py2.py3-none-any.whl (9.6 kB)
Collecting pyee<9.0.0,>=8.1.0
  Downloading pyee-8.2.2-py2.py3-none-any.whl (12 kB)
Requirement already satisfied: zipp>=0.5 in c:\users\pc-dell\anaconda3\envs\tf_gpu\lib\sit
e-packages (from importlib-metadata>=1.4->pyppeteer) (3.7.0)
Requirement already satisfied: colorama in c:\users\pc-dell\appdata\roaming\python\python3
10\site-packages (from tqdm<5.0.0,>=4.42.1->pyppeteer) (0.4.4)
Installing collected packages: pyee, appdirs, websockets, certifi, pyppeteer
  Attempting uninstall: certifi
    Found existing installation: certifi 2020.6.20
    Uninstalling certifi-2020.6.20:
      Successfully uninstalled certifi-2020.6.20
Successfully installed appdirs-1.4.4 certifi-2022.9.24 pyee-8.2.2 pyppeteer-1.0.2 websocke
ts-10.4
[notice] A new release of pip available: 22.2.2 -> 22.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```