

In [1]:



```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
```

In [2]:



```
1 data = pd.read_csv(r'netflix_titles.csv')  
2 data
```

Out[2]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006

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

In [3]:
8807 rows × 12 columns

```
1 data.info()
```

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

In [4]:

```
1 data.isnull().any()
```

Out[4]:

```
show_id      False
type         False
title        False
director     True
cast         True
country      True
date_added   True
release_year False
rating       True
duration     True
listed_in    False
description  False
dtype: bool
```

Task 1: Check for duplicacy, there is no duplicay

In [5]:

1 data.duplicated().any()

Out[5]:
False

conclusion Task 1: There is no duplicacy in data

Q. 1) For 'House of Cards', what is the Show Id and Who is the Director of this show ?

In [6]:

1 data[data['title']=='House of Cards']

Out[6]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	d
1059	s1060	TV Show	House of Cards	NaN	Kevin Spacey, Robin Wright, Kate Mara, Corey S...	United States	April 14, 2021	2018	TV-MA	s

Conclusion: Director is not specified in dataset, but director is James Foley (source: google)

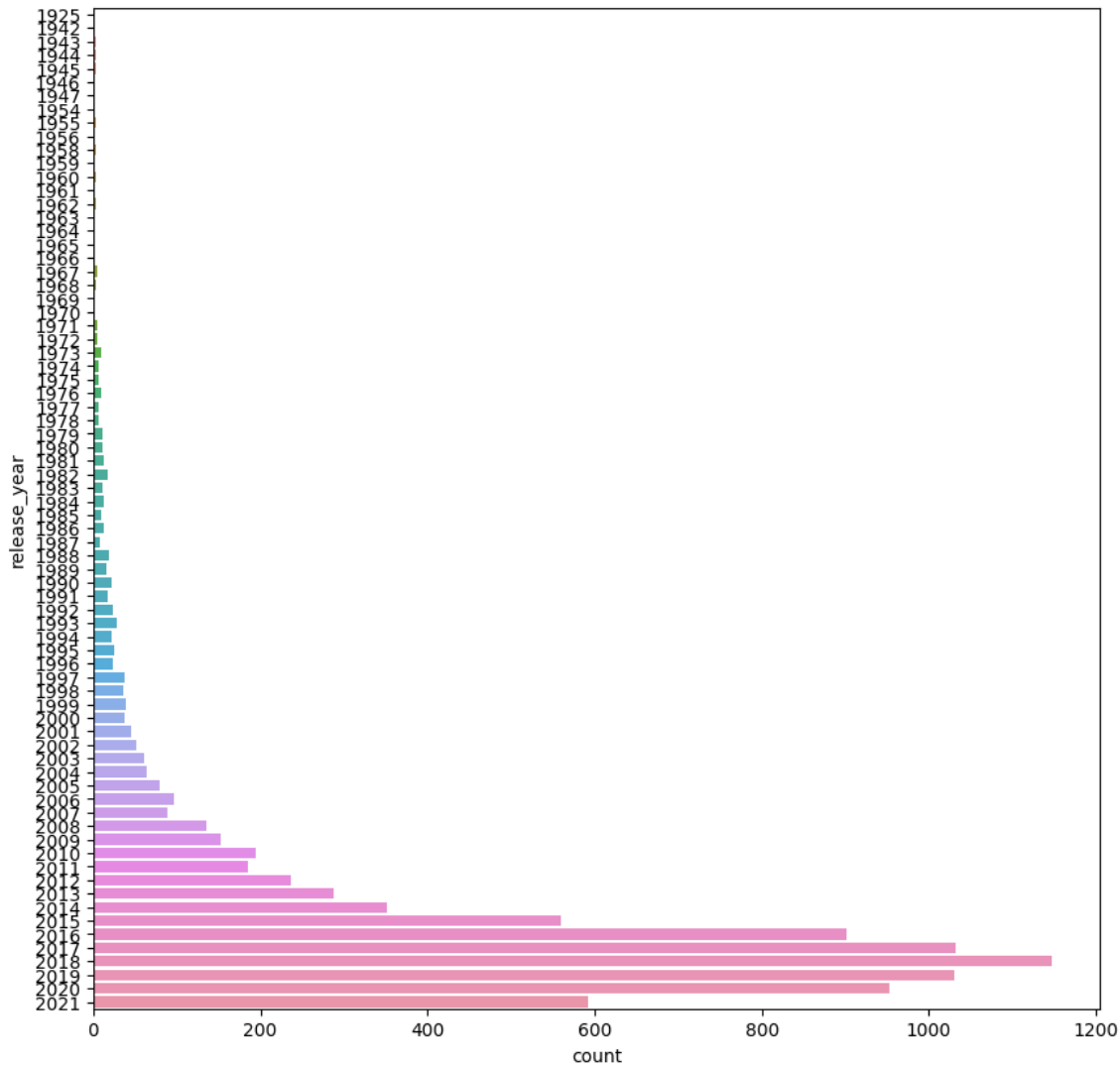
Q. 2) In which year the highest number of TV Shows & Movies were released? Show with Bar Graph

In [7]:

```
1 fig, ax = plt.subplots(figsize=(10, 10))
2 sns.countplot(y = 'release_year', data = data, ax = ax)
```

Out[7]:

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



conclusion: in 2018 the highest number of TV Shows & Movies were released

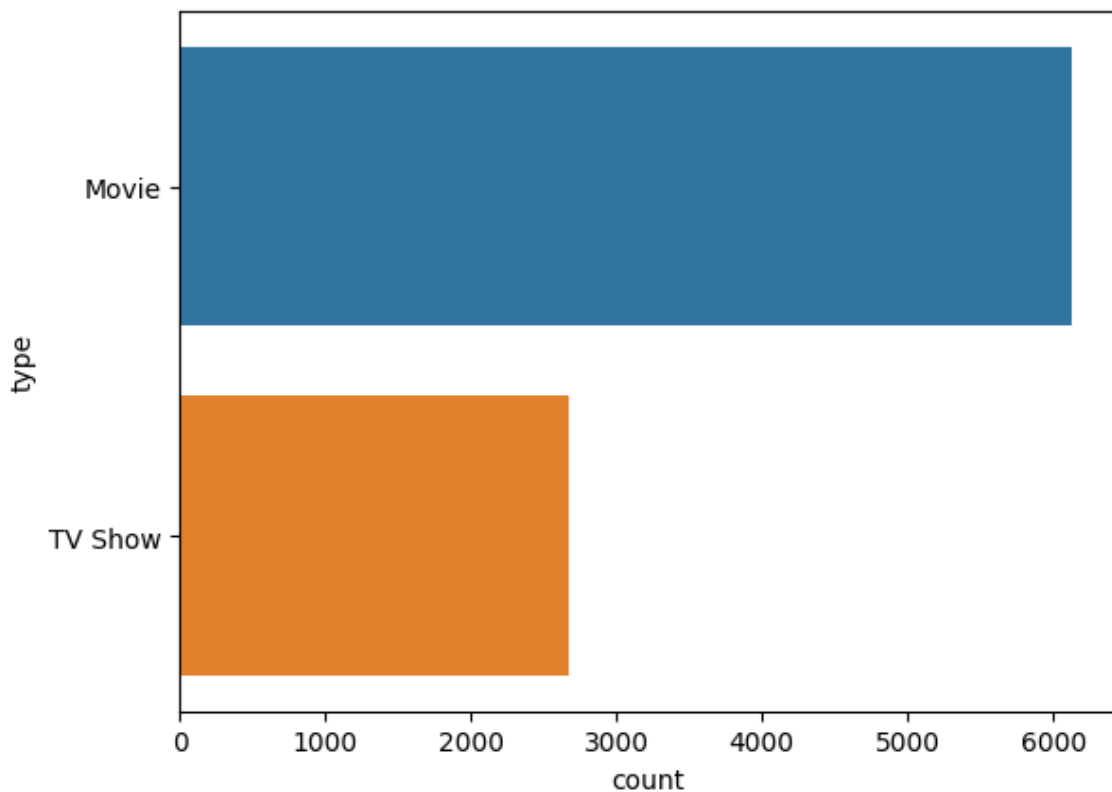
Q. 3) How many Movies & TV Shows are in the dataset ? Show with Bar Graph.

In [8]:

```
1 sns.countplot(y = 'type',data = data)
2 data['type'].value_counts()
```

Out[8]:

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



conclusion: 6131 movies and 2676 TV show

Q. 4) Show all the Movies that were released in the year 2000.



In [9]:



```
1 data[data['release_year']==2000][['title','type']]
```


Out[9]:

	title	type
180	The Nutty Professor II: The Klumps	Movie
350	Space Cowboys	Movie
359	The Original Kings of Comedy	Movie
566	Charlie's Angels	Movie
593	Snow Day	Movie
779	Battlefield Earth	Movie
952	The Whole Nine Yards	Movie
2173	Moesha	TV Show
3472	Rugrats in Paris: The Movie	Movie
4545	Monty Python: Before the Flying Circus	Movie
4723	Fiza	Movie
4956	Phir Bhi Dil Hai Hindustani	Movie
5098	Pokémon: Indigo League	TV Show
5245	Star Trek: Voyager	TV Show
5953	Joseph: King of Dreams	Movie
6006	28 Days	Movie
6147	American Psycho	Movie
6154	An American Tail: The Treasures of Manhattan I...	Movie
6311	Billy Elliot	Movie
6437	Center Stage	Movie
6440	Chal Mere Bhai	Movie
6543	Crouching Tiger, Hidden Dragon	Movie
6643	Dragonheart: A New Beginning	Movie
6753	Final Destination	Movie
6912	Hamara Dil Aapke Paas Hai	Movie
7013	How the Grinch Stole Christmas	Movie
7247	Kya Kehna	Movie
7316	Little Nicky	Movie
7702	Papa the Great	Movie
7766	Power Rangers Lightspeed Rescue	TV Show
7801	Pukar	Movie
7955	Scary Movie	Movie
7962	Scream 3	Movie
8196	The Art of War	Movie
8305	The Flintstones in Viva Rock Vegas	Movie
8722	What Lies Beneath	Movie

title

type

councclusion: above is the list of all the movies and TV showa released in 2000

8729Where the Money IsMovie

Q. 5) Show only the Titles of all TV Shows that were released in India only.

In [10]:

```
1 data[data['country']=='India'][['title','type']]
```

Out[10]:

	title	type
4	Kota Factory	TV Show
24	Jeans	Movie
39	Chhota Bheem	TV Show
50	Dharmakshetra	TV Show
66	Raja Rasoi Aur Anya Kahaniyan	TV Show
...
8773	Yanda Kartavya Aahe	Movie
8775	Yeh Meri Family	TV Show
8798	Zed Plus	Movie
8799	Zenda	Movie
8806	Zubaan	Movie

972 rows × 2 columns

conclusion: Above is the list of all the movies and tv shows released in india

In [11]:

```
1 a = data['director'].unique()
2 len(a)
```

Out[11]:

4529

Q. 6) Show Top 10 Directors, who gave the highest number of TV Shows & Movies to Netflix ?

In [55]:

```
1 # sns.countplot(x = 'director', data = data)
2 ndata = data.groupby('type')
3 ndata.head()
4 data['director'].value_counts()[:10]
```

Out[55]:

Rajiv Chilaka	19
Raúl Campos, Jan Suter	18
Marcus Raboy	16
Suhas Kadav	16
Jay Karas	14
Cathy Garcia-Molina	13
Jay Chapman	12
Youssef Chahine	12
Martin Scorsese	12
Steven Spielberg	11

Name: director, dtype: int64

Conclusion: Here is the list of top 10 directors having highest no. of TV Shows and movies

Q. 7) Show all the Records, where "Category is Movie and Type is Comedies" or "Country is United Kingdom".

In [13]:

```
1 filtered_data = data[((data['type'] == 'Movie') & (data['listed_in'].str.contains('C
```

In [14]:



1	filtered_data
---	---------------

Out[14]:

	show_id	type	title	director	cast	country	date_added	release_year
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021
13	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Klara Castanho, Lucca Picon, Júlia Gomes, Marc...	NaN	September 22, 2021	2021
22	s23	Movie	Avvai Shanmughi	K.S. Ravikumar	Kamal Hassan, Meena, Gemini Ganesan, Heera Raj...	NaN	September 21, 2021	1996
24	s25	Movie	Jeans	S. Shankar	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...	India	September 21, 2021	1998
...
8798	s8799	Movie	Zed Plus	Chandra Prakash Dwivedi	Adil Hussain, Mona Singh, K.K. Raina, Sanjay M...	India	December 31, 2019	2014
8799	s8800	Movie	Zenda	Avadhoot Gupte	Santosh Juvekar, Siddharth Chandekar, Sachit P...	India	February 15, 2018	2009
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006

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

Q. 8) In how many movies/shows, Tom Cruise was cast ?

In [15]:

```
1 data['cast'].str.contains('Tom Cruise')
```

Out[15]:

```
0      NaN
1     False
2     False
3      NaN
4     False
...
8802    False
8803     NaN
8804    False
8805    False
8806    False
Name: cast, Length: 8807, dtype: object
```

In [16]:

```
1 data['cast'].fillna('no data', inplace = True)
```

In [17]:

```
1 data['cast'].str.contains('Tom Cruise')
```

Out[17]:

```
0     False
1     False
2     False
3     False
4     False
...
8802    False
8803    False
8804    False
8805    False
8806    False
Name: cast, Length: 8807, dtype: bool
```

In [18]:

```
1 data[data['cast'].str.contains('Tom Cruise')]
```

Out[18]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
341	s342	Movie	Magnolia	Paul Thomas Anderson	John C. Reilly, Philip Baker Hall, Tom Cruise,...	United States	August 1, 2021	1999	
1254	s1255	Movie	Rain Man	Barry Levinson	Dustin Hoffman, Tom Cruise, Valeria Golino, Ge...	United States	March 1, 2021	1988	

conclusion : 2 movies in the dataset having tom cruise as actor

Q. 9) What are the different Ratings defined by Netflix ?

▪ Q. 9.1) How many Movies got the 'TV-14' rating, in Canada ?

▪ Q. 9.2) How many TV Shows got the 'R' rating, after year 2018 ?

In [19]:

```
1 data['rating'].unique()
```

Out[19]:

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

In [20]:



```
1 data[(data['rating']=='TV-14')&(data['country']=='Canada')].count()
```

Out[20]:

```
show_id      26
type         26
title        26
director     14
cast         26
country      26
date_added   26
release_year  26
rating       26
duration     26
listed_in    26
description   26
dtype: int64
```

In [21]:



```
1 data[(data['rating']=='R')&(data['release_year']>2018)].count()
```

Out[21]:

```
show_id      108
type         108
title        108
director     105
cast         108
country      101
date_added   108
release_year  108
rating       108
duration     108
listed_in    108
description   108
dtype: int64
```


Q. 10) What is the maximum duration of a Movie/Show on Netflix ?

In [22]:



```
1 movies = data[data['type']=='Movie']
2 movies['duration'].fillna('0 min',inplace = True)
3 df = movies['duration'].str.split(' ', expand=True)[0]
4 df = df.astype(int)
5 df
6 df.sort_values(ascending = True)
7
```

C:\Users\Parn\AppData\Local\Temp\ipykernel_13120\4137202063.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
movies['duration'].fillna('0 min',inplace = True)
```

Out[22]:

```
5541      0
5794      0
5813      0
3777      3
2713      5
...
2484     233
2487     237
2491     253
717      273
4253     312
```

Name: 0, Length: 6131, dtype: int32

In [23]:



```
1 tvshow = data[data['type']=='TV Show']
2 df2 = tvshow['duration'].str.split(' ',expand = True)[0]
3 df2.isna().any()
4 df2.astype(int)
5 df2.sort_values(ascending = True)
```

Out[23]:

```
3695    1
3539    1
3534    1
3527    1
3526    1
..
6504    9
8442    9
67      9
5803    9
1998    9
```

Name: 0, Length: 2676, dtype: object

conclusion: Max duration of Movies is 312 minutes & Max duration of TV show is 9 Series

• Q. 11) Which individual country has the Highest No. of TV Shows ?

In [44]:



```
1 data2 = data
2 data2 = data2[data2['type']=='TV Show']
3 data2['country'] = data2['country'].str.split(',')
4 f = f.explode('country')
5 data2 = data2.explode('country')
6 data2['country'].value_counts()
7 # United States have highest no of TV Shows
```

C:\Users\Parn\AppData\Local\Temp\ipykernel_13120\1696081142.py:3: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
data2['country'] = data2['country'].str.split(',')
```

Out[44]:

United States	847
United Kingdom	246
Japan	174
South Korea	164
United States	91
...	
Senegal	1
Luxembourg	1
Uruguay	1
United Arab Emirates	1
Switzerland	1

Name: country, Length: 102, dtype: int64

12) How can we sort the dataset by Year ?

In [46]:



```
1 data = data.sort_values(by = ['release_year'])
2 data
```

Out[46]:

	show_id	type	title	director	cast	country	date_added	release_year
4250	s4251	TV Show	Pioneers: First Women Filmmakers*	NaN	no data	NaN	December 30, 2018	1925
7790	s7791	Movie	Prelude to War	Frank Capra	no data	United States	March 31, 2017	1942
8205	s8206	Movie	The Battle of Midway	John Ford	Henry Fonda, Jane Darwell	United States	March 31, 2017	1942
8660	s8661	Movie	Undercover: How to Operate Behind Enemy Lines	John Ford	no data	United States	March 31, 2017	1943
8739	s8740	Movie	Why We Fight: The Battle of Russia	Frank Capra, Anatole Litvak	no data	United States	March 31, 2017	1943
...
862	s863	TV Show	Special	NaN	Ryan O'Connell, Jessica Hecht, Punam Patel, Ma...	United States	May 20, 2021	2021
861	s862	Movie	Little Singham Future mein Satakli	Prakash Satam	Anamaya Verma, Ganesh Divekar, Jigna Bharadhma...	NaN	May 20, 2021	2021
857	s858	TV Show	The Neighbor	NaN	Quim Gutiérrez, Clara Lago, Adrián Pino, Catel...	Spain	May 21, 2021	2021
876	s877	TV Show	Alma Matters	NaN	no data	India	May 14, 2021	2021
965	s966	TV Show	Yasuke	NaN	LaKeith Stanfield, Takehiro Hira, Maya Tanida,...	NaN	April 29, 2021	2021

Q. 13) Find all the instances where: Category is 'Movie' and Type is 'Dramas' or Category is 'TV Show' & Type is 'Kids' TV'.**

Task. 1) Is there any Duplicate Record in this dataset? If yes, then remove the duplicate records."Country is United Kingdom"

In [50]:

```
1 data[(((data['type']=='Movie') & (data['listed_in'].str.contains('Dramas')))|((data[
2 data.duplicated().any()
3 # no duplicate data
```

Out[50]:

False