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

Defining Problem Statement and Analysing basic metrics

We have to analize the data and give recommendation which type of movies and tv shows should be produced by netflix or should be streamed in netflix with the help of basic metrics like group by and different types of plots

In [2]:
```
netflix_data = pd.read_csv('netflix.csv')
```

In [3]:
```
netflix_data.head()
```

Out[3]:

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

The first step is to unnesting the data in the columns

In [4]:
```
netflix_data['director'] = netflix_data['director'].str.split(', ')
netflix_data['cast'] = netflix_data['cast'].str.split(', ')
netflix_data['country'] = netflix_data['country'].str.split(', ')
netflix_data['listed_in']  = netflix_data['listed_in'].str.split(', ')
```

Here I have used the explode function inorder to unnest the data in the columns

In [5]:
```
new_netflix_data = netflix_data.explode('director')
new_netflix_data = new_netflix_data.explode('cast')
new_netflix_data = new_netflix_data.explode('country')
new_netflix_data = new_netflix_data.explode('listed_in')
```

In [6]:
```
new_netflix_data.head()
```

Out[6]:

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

In [7]:
```
new_netflix_data.shape
```

Out[7]:
```
(201991, 12)
```

from th eshape of the data it is clear that their are 201991 rows or data points or entries and 12 columns or features

In [8]:
```
new_netflix_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 201991 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count   Dtype 
---  ------        --------------   ----- 
 0   show_id       201991 non-null  object
 1   type          201991 non-null  object
 2   title         201991 non-null  object
 3   director      151348 non-null  object
 4   cast          199845 non-null  object
 5   country       190094 non-null  object
 6   date_added    201833 non-null  object
 7   release_year  201991 non-null  int64 
 8   rating        201924 non-null  object
 9   duration      201988 non-null  object
 10  listed_in     201991 non-null  object
 11  description   201991 non-null  object
dtypes: int64(1), object(11)
memory usage: 20.0+ MB
```

from the above info it is clear that except release_year every other feature is string data type and the release_year is int data type

In [9]:
```
new_netflix_data.isna().sum()
```

Out[9]:
```
show_id            0
type               0
title              0
director       50643
cast            2146
country        11897
date_added       158
release_year       0
rating            67
duration           3
listed_in          0
description        0
dtype: int64
```

From the above cell it is clear that in the directors,cast,country,data_added,rating and duration features are having NAN values

In [10]:
```
new_netflix_data.describe()
```

Out[10]:

	release_year
count	201991.000000
mean	2013.452891
std	9.003933
min	1925.000000
25%	2012.000000
50%	2016.000000
75%	2019.000000
max	2021.000000

the starting year is 1925 the ending year is 2021 most of the movies are released between the years 2012 and 2021 as the difference between the mean and the median is not that great so their is no outlyer which is effecting the mean value

In [11]:
```
new_netflix_data.describe(include='object')
```

Out[11]:

	show_id	type	title	director	cast	country	date_added	rating	duration	liste
count	201991	201991	201991	151348	199845	190094	201833	201924	201988	20
unique	8807	2	8807	4993	36439	127	1767	17	220	
top	s7165	Movie	Kahlil Gibran's The Prophet	Martin Scorsese	Liam Neeson	United States	January 1, 2020	TV-MA	1 Season	Dra
freq	700	145843	700	419	161	59349	3730	73867	35035	2

The Movie occurs 145843 times US has the most movies and tv shows combbined 59349 most of the movies are added on jan 1 2020 3730 TV-MA is the top rating

In [12]:
```
movie_netflix_data = new_netflix_data[new_netflix_data['type']=='Movie']
```

Here we generally dividing the data in to 2 data frames 1st for Movies and the 2nd for Tv Showsso that the analysis will be easy to perform

In [13]:
```
movie_netflix_data.head()
```

Out[13]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	durat
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 r
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	NaN	September 24, 2021	2021	PG	91 r
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	NaN	September 24, 2021	2021	PG	91 r
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	James Marsden	NaN	September 24, 2021	2021	PG	91 r
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Sofia Carson	NaN	September 24, 2021	2021	PG	91 r

In [14]:
```
movie_netflix_data.isna().sum()
```

Out[14]:
```
show_id          0
type             0
title            0
director      1285
cast          1328
country       6199
date_added       0
release_year     0
rating           9
duration         3
listed_in        0
description      0
dtype: int64
```

In [15]:
```
tv_show_netflix_data = new_netflix_data[new_netflix_data['type']=='TV Show']
```

In [16]:
```
tv_show_netflix_data.isna().sum()
```

Out[16]:
```
show_id            0
type               0
title              0
director       49358
cast             818
country         5698
date_added       158
release_year       0
rating            58
duration           0
listed_in          0
description        0
dtype: int64
```

Movie Recommendation Starts

In [17]:
```
movie_netflix_data['cast'].value_counts()
```

Out[17]:
```
Liam Neeson        161
Alfred Molina      157
John Krasinski     138
Salma Hayek        130
Frank Langella     128
                  ...
Sneha Koorse         1
Will Ryan            1
Steve Rosenbaum      1
Nina Richmond        1
Jason Beghe          1
Name: cast, Length: 25951, dtype: int64
```

here the Most frequently occuring actor is Liam Neeson followed by Alfred Molina

In [18]:
```
movie_netflix_data['director'].value_counts()
```

Out[18]:
```
Martin Scorsese        419
Youssef Chahine        409
Cathy Garcia-Molina    356
Steven Spielberg       355
Lars von Trier         336
                      ...
Halder Gomes             1
Mike Rossiter            1
Jerry Kolber             1
Catrin Einhorn           1
Brandon Jones            1
Name: director, Length: 4777, dtype: int64
```

Mosst frequent director is Martin Scorsese followed by Youssef Chahine

In [19]:
```
movie_netflix_data = movie_netflix_data.dropna()
```

Droping the NAN columns (droping is the treatment which I chose)

In [20]:
```
movie_netflix_data
```

Out[20]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	du
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	1
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	1
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	1
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	Ghana	September 24, 2021	1993	TV-MA	1
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	Ghana	September 24, 2021	1993	TV-MA	1
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	1
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	1
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	1
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	1
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	1

137532 rows × 12 columns

In [21]:
```
movie_netflix_data.isna().sum()
```

Out[21]:
```
show_id         0
type            0
title           0
director        0
cast            0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
description     0
dtype: int64
```

The is to make sure that NAN values are droped

In [22]:
```
movie_netflix_data['duration'] = movie_netflix_data['duration'].apply(lambda x:ir
```

Here I am extracting the integer part of the duration using(lambda function) in side apply

In [23]:
```
movie_netflix_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 137532 entries, 7 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count   Dtype
---  ------        --------------   -----
 0   show_id       137532 non-null  object
 1   type          137532 non-null  object
 2   title         137532 non-null  object
 3   director      137532 non-null  object
 4   cast          137532 non-null  object
 5   country       137532 non-null  object
 6   date_added    137532 non-null  object
 7   release_year  137532 non-null  int64
 8   rating        137532 non-null  object
 9   duration      137532 non-null  int64
 10  listed_in     137532 non-null  object
 11  description   137532 non-null  object
dtypes: int64(2), object(10)
memory usage: 13.6+ MB
```

now the int type data structures are 2 columns (release_year and duration)

In [24]:
```
netflix_data = movie_netflix_data[pd.to_datetime(movie_netflix_data['date_added'])
```

Here I am removing the improper data such as the include year is before the release_year

In [25]:
```
movie_netflix_date = movie_netflix_data.drop(columns='description',inplace=True)
```

```
C:\Users\satis\anaconda3\lib\site-packages\pandas\core\frame.py:3997: SettingWi
thCopyWarning:
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/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pyd
ata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c
opy)
  errors=errors,
```

Droping the description column

In [26]:
```
movie_netflix_data
```

Out[26]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	du
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	Ghana	September 24, 2021	1993	TV-MA	
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	Ghana	September 24, 2021	1993	TV-MA	
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	

137491 rows × 11 columns

In [100]:
```
movie_netflix_data.reset_index().reset_index().drop(columns='index',inplace=True)
```

In [101]:
```
movie_netflix_data
```

Out[101]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	du
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	Ghana	September 24, 2021	1993	TV-MA	
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	Ghana	September 24, 2021	1993	TV-MA	
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	

137436 rows × 11 columns

In [30]:
```
movie_netflix_data.groupby('director')['cast'].count().sort_values(ascending=Fals
```

Out[30]:
```
director
Martin Scorsese         419
Youssef Chahine         409
Cathy Garcia-Molina     356
Steven Spielberg        355
Lars von Trier          336
                       ...
Joana Mazzucchelli        1
Jim Gable                 1
Jessica Yu                1
Natalia Valdebenito       1
Richard Mears             1
Name: cast, Length: 4165, dtype: int64
```

In [31]:
```
movie_netflix_data.drop_duplicates(keep='first',inplace=True)
```

```
C:\Users\satis\anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWith
CopyWarning:
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/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pyd
ata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c
opy)
  """Entry point for launching an IPython kernel.
```

Droping the duplicates and keeping the first appiring values

In [32]:
```
movie_netflix_data['director'].value_counts()
```

Out[32]:
```
Martin Scorsese         419
Youssef Chahine         409
Cathy Garcia-Molina     356
Steven Spielberg        355
Lars von Trier          336
                       ...
Richard Miron             1
Shawn Rech                1
Matthew Cooke             1
Will Lovelace             1
Leslie Iwerks             1
Name: director, Length: 4165, dtype: int64
```

In [33]:
```
movie_netflix_data.head()
```

Out[33]:

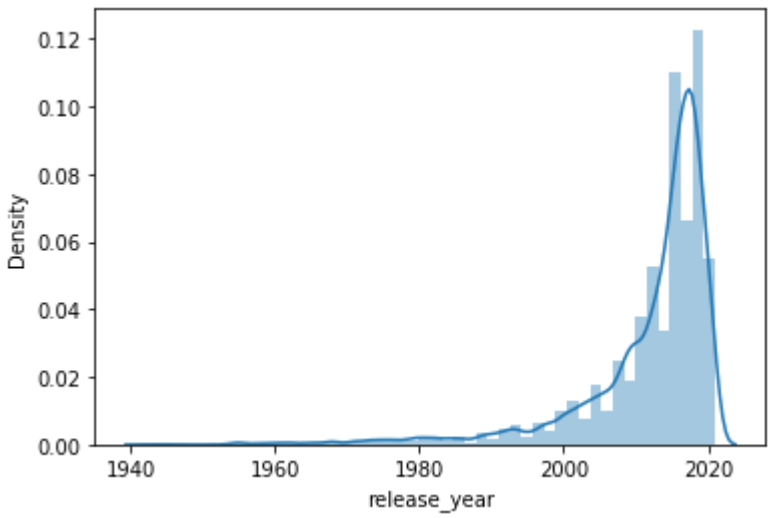
	show_id	type	title	director	cast	country	date_added	release_year	rating	duration
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	125
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	125
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	United States	September 24, 2021	1993	TV-MA	125
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	Ghana	September 24, 2021	1993	TV-MA	125
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	Ghana	September 24, 2021	1993	TV-MA	125

In [71]:
```
movie_netflix_data['duration'].unique()
```

Out[71]:
```
array([125, 104, 127, 166, 103,  97, 106,  96, 124, 116,  98,  91, 115,
       122,  99,  88, 100, 102,  93,  95,  85,  83, 182, 147,  90, 128,
       143, 119, 114, 118, 108, 117, 121, 142, 113, 154, 120,  82,  94,
       109, 101, 105,  86, 229,  76,  89, 110, 156, 112, 129, 107, 135,
       136, 165, 150, 133, 145,  92,  64,  59, 111,  87, 148, 189, 141,
       130,  68, 131, 126, 155, 123,  84,  13,  77,  74,  49,  72,  78,
        70, 132, 140,  81, 138, 149,  15, 224, 162,  60,  65, 137,  75,
        32, 158, 164, 173, 181,  73,  24,  80, 139, 151,  22, 134,  58,
        52,  71, 161,  53,   8,  46,  57,  28,  66,  50,  79,  26,  54,
        48,  45, 171,  42,  27,  51,  69,  47,  33,  44,  29, 146,  61,
        63, 157, 203,  25,  30,  62, 194,  55, 177, 237, 195, 253, 152,
        67, 190, 160, 208, 180, 144, 174, 170, 192, 209, 187, 185, 172,
       186, 193, 176,  17,  56, 169,  40,  20,  12, 168, 153, 159, 214,
        31, 163,  14, 179,  38,  23,  43, 200, 196, 167,  41,  37,  35,
       178, 228,  18, 205, 201, 191], dtype=int64)
```

In [107]:
```
ax = sns.distplot(movie_netflix_data['release_year'])
plt.show()
```

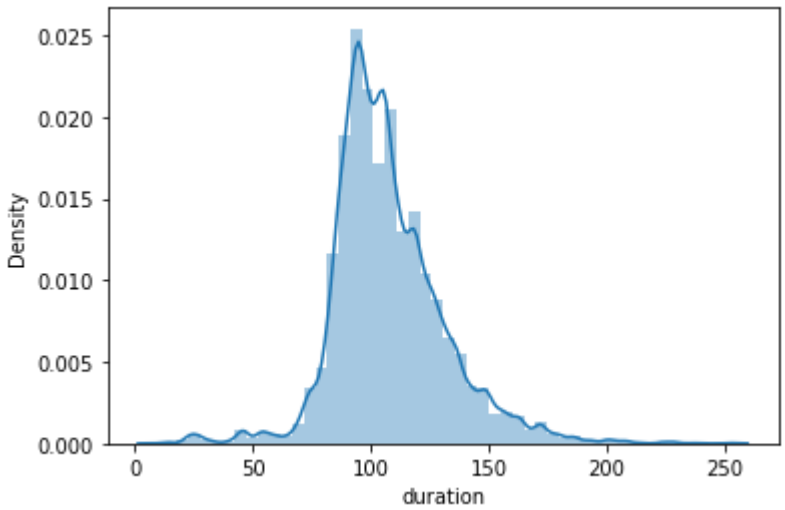
```
C:\Users\satis\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



Here we can see that movies releasing started increasing from the year 2000 and keep on increasing and decreased slightly

In [50]:
```
sns.distplot(movie_netflix_data['duration'])
plt.show()
```

```
C:\Users\satis\anaconda3\lib\site-packages\seaborn\distributions.py:2619: Futur
eWarning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level function
with similar flexibility) or `histplot` (an axes-level function for histogram
s).
  warnings.warn(msg, FutureWarning)
```

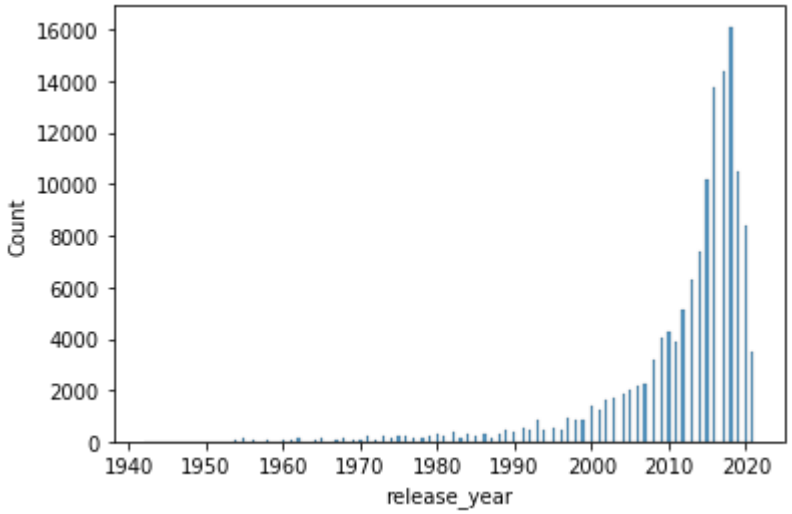


most of the movies are around 100 minutes movies

In [125]:
```
# ax = sns.countplot(y='release_year',data=movie_netflix_data)
# plt.show()
```

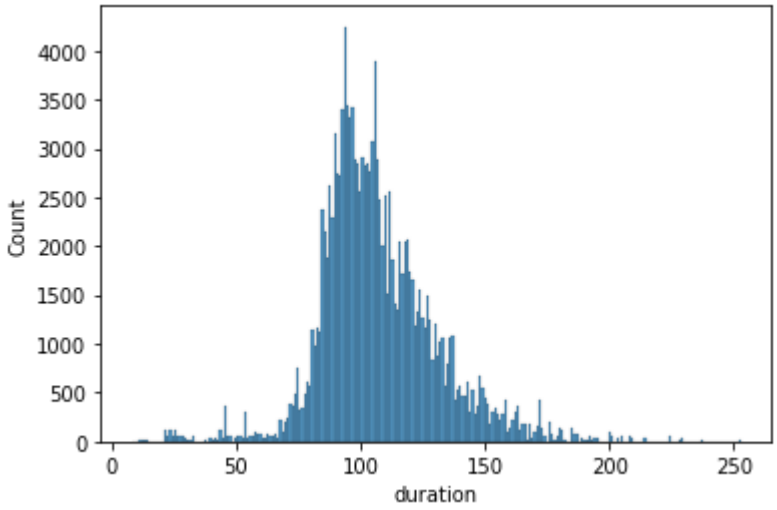
In [126]:
```
# ax = sns.countplot(y='duration',data=movie_netflix_data)
# plt.show()
```

In [58]:
```
ax = sns.histplot(movie_netflix_data['release_year'])
plt.show()
```

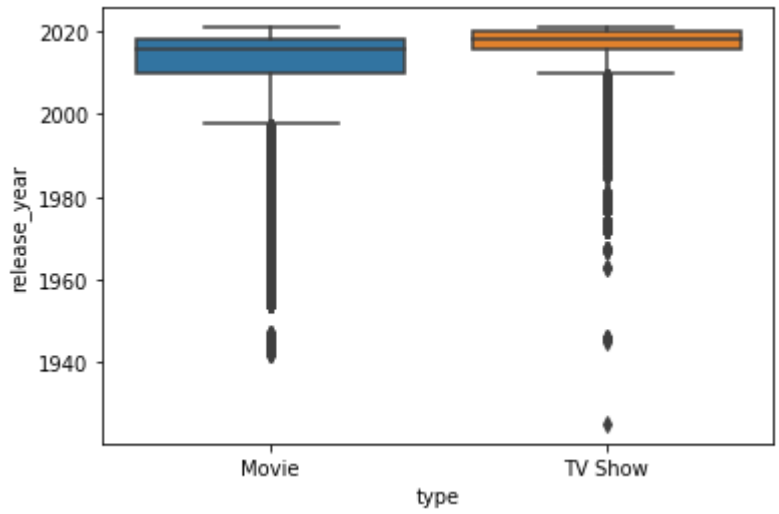


From the histplot also it is confirmed that movies started increasing around the year 2000

In [60]:
```
ax = sns.histplot(movie_netflix_data['duration'])
plt.show()
```

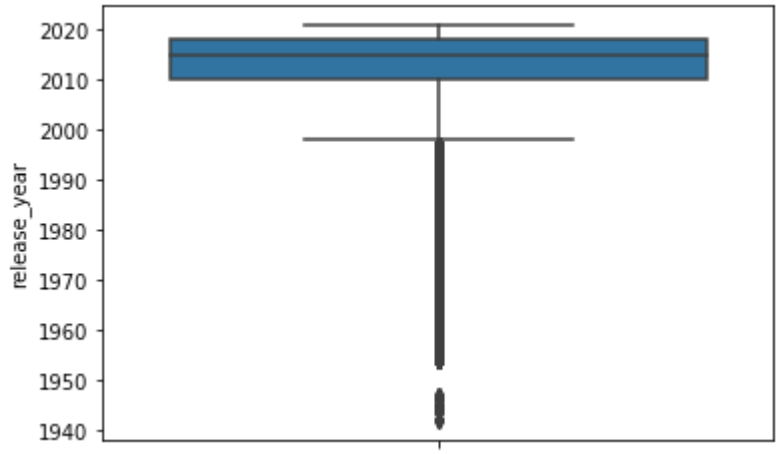


In [81]:
```
ax = sns.boxplot(x = 'type',y = 'release_year',data=new_netflix_data)
plt.show()
```

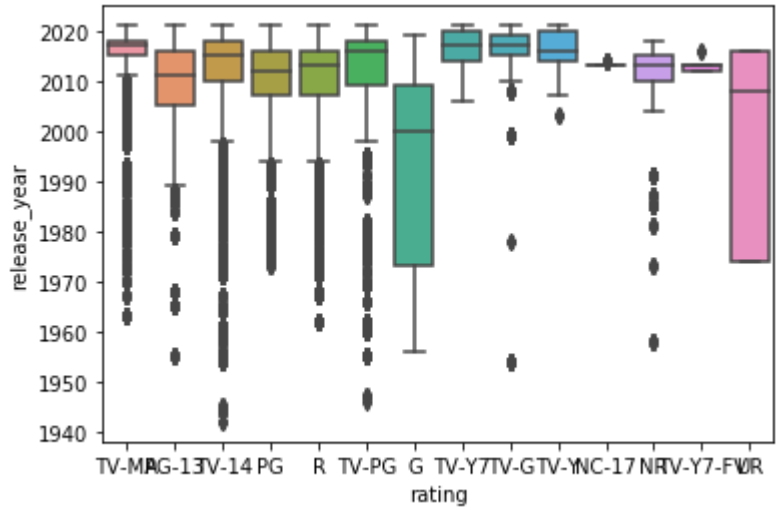


From the above box plot it is clear that at the high period of movies Tv shows started sslowly and increased gradually and surpassed moivies

In [73]:
```
ax = sns.boxplot(y = 'release_year',data=movie_netflix_data)
plt.show()
```

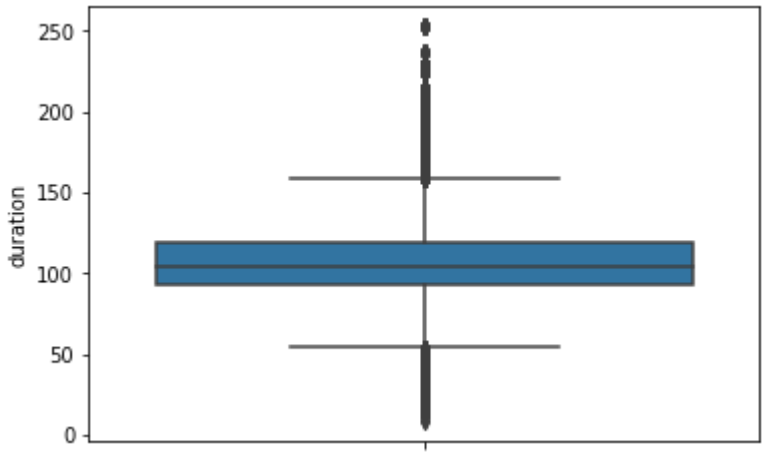


In [123]:
```
ax = sns.boxplot(x = 'rating',y = 'release_year',data=movie_netflix_data)
plt.show()
```



outlier are present in all of the ratings except G ,Y7and TV-R

In [74]:
```
ax = sns.boxplot(y='duration',data=movie_netflix_data)
```

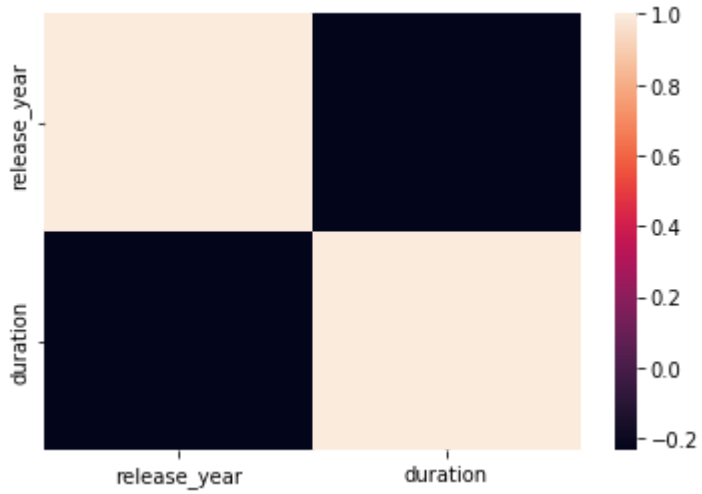


In [75]:
```
movie_netflix_data.corr()
```

Out[75]:

	release_year	duration
release_year	1.000000	-0.231447
duration	-0.231447	1.000000

In [80]:
```
ax = sns.heatmap(movie_netflix_data.corr())
```



In [ ]:
```
```

In [34]:
```
movie_netflix_data.groupby('director')['cast'].max().reset_index()
```

Out[34]:

	director	cast
0	A. L. Vijay	Yogi Babu
1	A. Raajdheep	Yogi Babu
2	A. Salaam	Sulakshana Pandit
3	A.R. Murugadoss	S.J. Surya
4	Aadish Keluskar	Rohit Kokate
...	...	...
4160	Éric Warin	Terrence Scammell
4161	Ísold Uggadóttir	Patrik Nökkvi Pétursson
4162	Óskar Thór Axelsson	Þröstur Leó Gunnarsson
4163	Ömer Faruk Sorak	Şafak Sezer
4164	Şenol Sönmez	Özgür Emre Yıldırım

4165 rows × 2 columns

The output of the above cell shows the favourite actor of that directors

In [35]:
```
movie_netflix_data.groupby('country')['cast'].max().reset_index()
```

Out[35]:

	country	cast
0		Walid Abdul Salam
1	Afghanistan	Sohrab Nazari
2	Albania	Marco Giallini
3	Algeria	Youssef Wahby
4	Angola	Raul Rosario
...	...	...
108	Vatican City	Pope Francis
109	Venezuela	Victoria Raposo
110	Vietnam	Truong Hoang Hanh Thy
111	West Germany	Zygmunt Zintel
112	Zimbabwe	Zihlo

113 rows × 2 columns

The output of the above cell is country specific best actor(in terms of most number of movies acted in)

In [36]:
```
movie_netflix_data.groupby('listed_in')['director'].max().reset_index()
```

Out[36]:

	listed_in	director
0	Action & Adventure	Ömer Faruk Sorak
1	Anime Features	Zhao Ji
2	Children & Family Movies	Éric Warin
3	Classic Movies	Youssef Chahine
4	Comedies	Şenol Sönmez
5	Cult Movies	Walter Hill
6	Documentaries	Zatella Beatty
7	Dramas	Şenol Sönmez
8	Faith & Spirituality	Wim Wenders
9	Horror Movies	Óskar Thór Axelsson
10	Independent Movies	Ísold Uggadóttir
11	International Movies	Şenol Sönmez
12	LGBTQ Movies	Ángeles Reiné
13	Movies	Wayne Orr
14	Music & Musicals	Álex de la Iglesia
15	Romantic Movies	Şenol Sönmez
16	Sci-Fi & Fantasy	Àlex Pastor
17	Sports Movies	Zatella Beatty
18	Stand-Up Comedy	Y. Joon Chung
19	Thrillers	Àlex Pastor

The output of the above cell is most no of movies directed by the director in that particular genre

In [37]:
```
movie_netflix_data.groupby('listed_in')['cast'].max().reset_index()
```

Out[37]:

	listed_in	cast
0	Action & Adventure	Şafak Sezer
1	Anime Features	Zhang He
2	Children & Family Movies	İlknur Külahlıoglu
3	Classic Movies	Čestmír Řanda
4	Comedies	Şinasi Yurtsever
5	Cult Movies	Zulay Henao
6	Documentaries	Ángel Mosqueda
7	Dramas	Şükrü Özyıldız
8	Faith & Spirituality	Zaskia Adya Mecca
9	Horror Movies	Ṣọpẹ́ Dìrísù
10	Independent Movies	Ṣọpẹ́ Dìrísù
11	International Movies	Şükrü Özyıldız
12	LGBTQ Movies	Érika Mader
13	Movies	k.d. lang
14	Music & Musicals	Ólafur Darri Ólafsson
15	Romantic Movies	Şinasi Yurtsever
16	Sci-Fi & Fantasy	Álvaro Roig
17	Sports Movies	Özlem Türkad
18	Stand-Up Comedy	Zach Galifianakis
19	Thrillers	İlker Kaleli

The output of the above cell is most no of movies acted by the actor in that particular genre

In [122]:
```
# movie_netflix_data.groupby('listed_in')['director'].max().reset_index()
```

In [39]:
```
movie_netflix_data.groupby('country')['listed_in'].max().reset_index()
```

Out[39]:

	country	listed_in
0		International Movies
1	Afghanistan	International Movies
2	Albania	International Movies
3	Algeria	International Movies
4	Angola	International Movies
...	...	...
108	Vatican City	International Movies
109	Venezuela	Thrillers
110	Vietnam	Thrillers
111	West Germany	Thrillers
112	Zimbabwe	Romantic Movies

113 rows × 2 columns

The output of the above cell shows which type of genre is most popolar in that particular country

In [40]:
```
movie_netflix_data.groupby('country')['director'].max().reset_index()
```

Out[40]:

	country	director
0		Najwa Najjar
1	Afghanistan	Pieter-Jan De Pue
2	Albania	Antonio Morabito
3	Algeria	Youssef Chahine
4	Angola	Maradona Dias Dos Santos
...	...	...
108	Vatican City	Wim Wenders
109	Venezuela	Sebastián Schindel
110	Vietnam	Victor Vu
111	West Germany	Mel Stuart
112	Zimbabwe	Tomas Brickhill

113 rows × 2 columns

The output of the above cell is most number of movies directed by the

director in his country

In [41]:
```
movie_netflix_data.groupby('director')['duration'].max().sort_values(ascending=Fa
```

Out[41]:

	director	duration
0	Houssam El-Din Mustafa	253
1	Samir Al Asfory	237
2	Sergio Leone	229
3	Raj Kapoor	228
4	Ashutosh Gowariker	224
...	...	...
4160	Harry Chaskin	14
4161	Mathieu Auvray	12
4162	Floyd Russ	12
4163	Timothy Ware-Hill	8
4164	Arnon Manor	8

4165 rows × 2 columns

Maximum length of movie directed by the director

This recomendatin is for the country Vietnam we can draw the similar conclusions for the above data to the remaining countries as well

for Example if we want to recommend a movie in the country Vietnam as per the data triller movies are most popular in that country so it is best to release triller movies in that country.

When comes to the director Victor Vu is the popular director so the movies directed by him are most likely to get outreacch in vietnam

When comes to the actor Truong Hoang Hanh Thy is the famous actor in the country so it is best to release his movies inn the vietnam.

Tv show Recommendation Starts

In [82]:
```
tv_show_netflix_data.head()
```

Out[82]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	Ir
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	.
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	
1	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	Ir
1	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	.

In [85]:
```
tv_show_netflix_data['cast'].value_counts()
```

Out[85]:
```
David Attenborough    82
Takahiro Sakurai      56
Yuki Kaji             45
Ai Kayano             41
Junichi Suwabe        39
                      ..
Andrea Savage          1
Tracee Ellis Ross      1
Mousam                 1
Alexa Alemanni         1
Shawtane Bowen         1
Name: cast, Length: 14863, dtype: int64
```

The highest number of Tv shows was acted by the acter David Attenborough followed by Takahiro Sakurai

In [89]:
```
tv_show_netflix_data = tv_show_netflix_data.dropna()
```

Drop the columns with the values NAN

In [90]:
```
tv_show_netflix_data
```

Out[90]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	du
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Sue Perkins	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Sue Perkins	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mary Berry	United Kingdom	September 24, 2021	2021	TV-14	Se
...	...	...	...	...	...	...	...	...	...	
8599	s8600	TV Show	Toast of London	Michael Cumming	Harry Peacock	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Harry Peacock	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Tracy Ann Oberman	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Tracy Ann Oberman	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Tracy Ann Oberman	United Kingdom	September 1, 2017	2015	TV-MA	Se

5560 rows × 12 columns

In [91]:
```
tv_show_netflix_data.isna().sum()
```

Out[91]:
```
show_id         0
type            0
title           0
director        0
cast            0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
description     0
dtype: int64
```

This is to confirm the droping of NAN values

In [92]:
```
tv_show_netflix_data = tv_show_netflix_data[pd.to_datetime(tv_show_netflix_data['
```

Removing the rows with data added is before the release date

In [93]:
```
tv_show_netflix_date = tv_show_netflix_data.drop(columns='description',inplace=Tr
```

```
C:\Users\satis\anaconda3\lib\site-packages\pandas\core\frame.py:3997: SettingWi
thCopyWarning:
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/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pyd
ata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c
opy)
  errors=errors,
```

Droping the column description

In [94]:
```
tv_show_netflix_data
```

Out[94]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	du
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Sue Perkins	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Sue Perkins	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mary Berry	United Kingdom	September 24, 2021	2021	TV-14	Se
...	...	...	...	...	...	...	...	...	...	
8599	s8600	TV Show	Toast of London	Michael Cumming	Harry Peacock	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Harry Peacock	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Tracy Ann Oberman	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Tracy Ann Oberman	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Tracy Ann Oberman	United Kingdom	September 1, 2017	2015	TV-MA	Se

5530 rows × 11 columns

In [97]:
```
tv_show_netflix_data.reset_index().drop(columns='index',inplace=True)
```

In [98]:
```
tv_show_netflix_data
```

Out[98]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	du
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Sue Perkins	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Sue Perkins	United Kingdom	September 24, 2021	2021	TV-14	Se
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mary Berry	United Kingdom	September 24, 2021	2021	TV-14	Se
...	...	...	...	...	...	...	...	...	...	
8599	s8600	TV Show	Toast of London	Michael Cumming	Harry Peacock	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Harry Peacock	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Tracy Ann Oberman	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Tracy Ann Oberman	United Kingdom	September 1, 2017	2015	TV-MA	Se
8599	s8600	TV Show	Toast of London	Michael Cumming	Tracy Ann Oberman	United Kingdom	September 1, 2017	2015	TV-MA	Se

5530 rows × 11 columns

In [102]:
```
tv_show_netflix_data.groupby('director')['cast'].count().sort_values(ascending=Fa
```

Out[102]:
```
director
Noam Murro          189
Thomas Astruc       160
Alan Poul           104
Houda Benyamina     104
Laïla Marrakchi     104
                   ...
Mick Grogan           2
Eric Abrams           1
Oliver Stone          1
Glenn Weiss           1
Michael Simon         1
Name: cast, Length: 190, dtype: int64
```

In [104]:
```
tv_show_netflix_data.drop_duplicates(keep='first',inplace=True)
```

```
C:\Users\satis\anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWith
CopyWarning:
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/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pyd
ata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c
opy)
  """Entry point for launching an IPython kernel.
```

In [105]:
```
tv_show_netflix_data['director'].value_counts()
```

Out[105]:
```
Noam Murro          189
Thomas Astruc       160
Houda Benyamina     104
Laïla Marrakchi     104
Damien Chazelle     104
                   ...
Tig Notaro            2
Oliver Stone          1
Eric Abrams           1
Glenn Weiss           1
Michael Simon         1
Name: director, Length: 190, dtype: int64
```

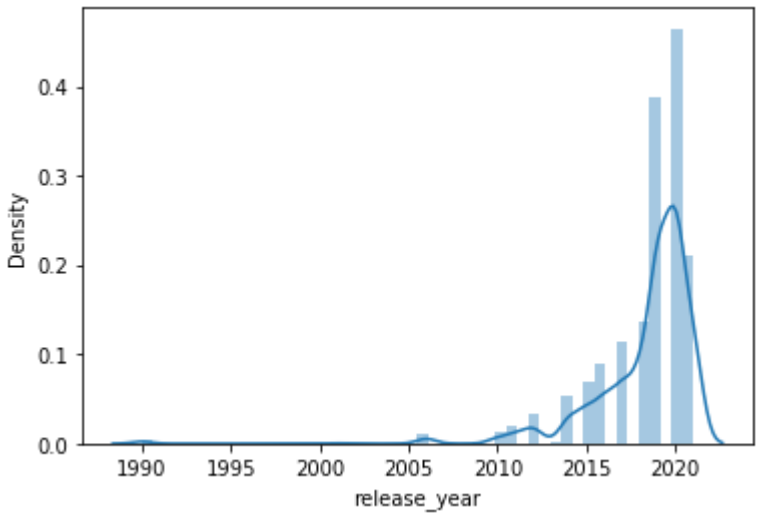
In [106]:
```
tv_show_netflix_data.head()
```

Out[106]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duratic
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc	United Kingdom	September 24, 2021	2021	TV-14	Seasor
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc	United Kingdom	September 24, 2021	2021	TV-14	Seasor
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Sue Perkins	United Kingdom	September 24, 2021	2021	TV-14	Seasor
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Sue Perkins	United Kingdom	September 24, 2021	2021	TV-14	Seasor
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mary Berry	United Kingdom	September 24, 2021	2021	TV-14	Seasor

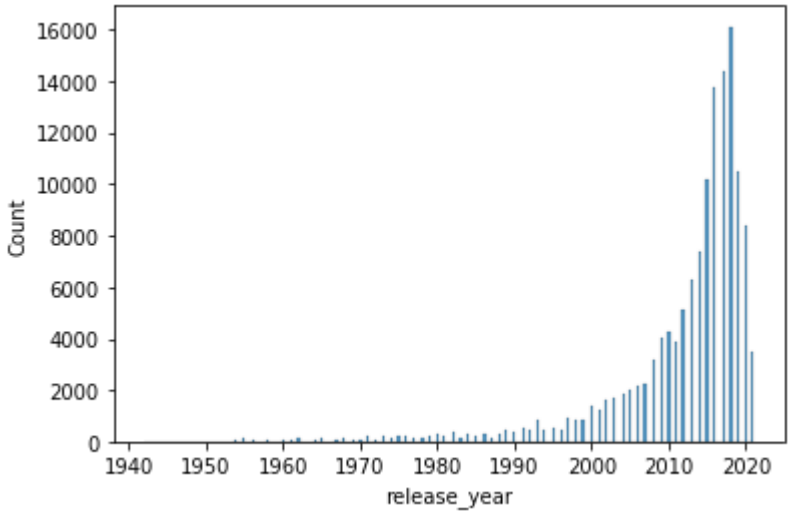
In [108]:
```
ax = sns.distplot(tv_show_netflix_data['release_year'])
plt.show()
```

```
C:\Users\satis\anaconda3\lib\site-packages\seaborn\distributions.py:2619: Futur
eWarning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level function
with similar flexibility) or `histplot` (an axes-level function for histogram
s).
  warnings.warn(msg, FutureWarning)
```



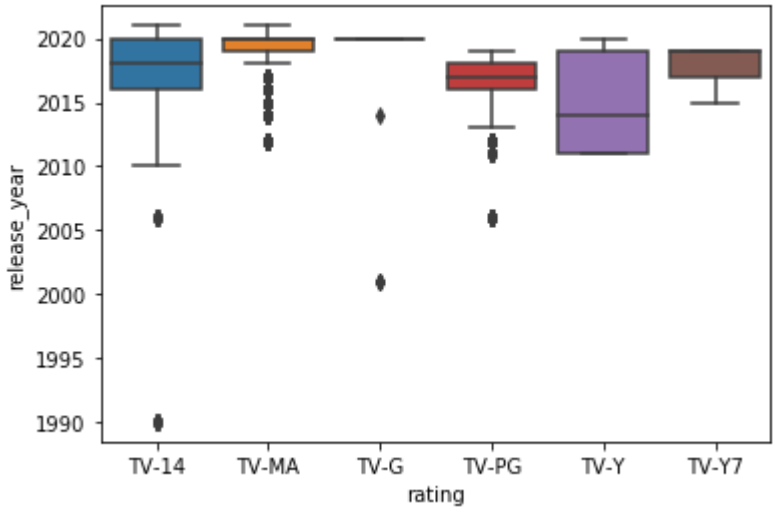
most no of Tv shows started increasing around 2010 and keeps on increasing from their on

In [110]:
```
ax = sns.histplot(movie_netflix_data['release_year'])
plt.show()
```



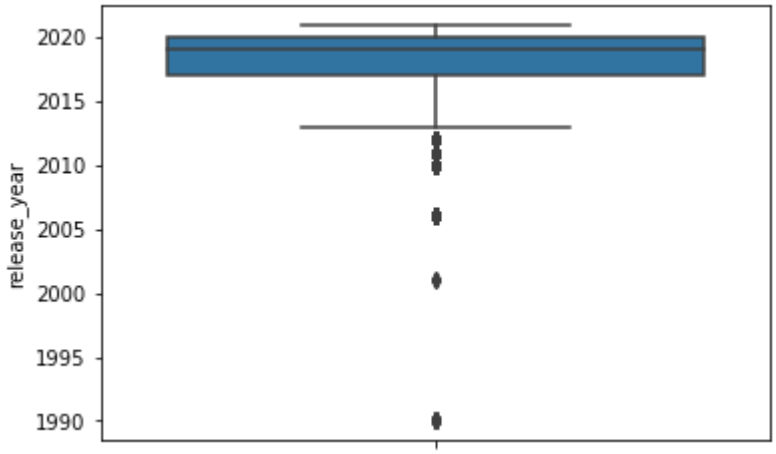
most no of Tv shows started increasing around 2010 and keeps on increasing from their on

In [111]:
```
ax = sns.boxplot(x = 'rating',y = 'release_year',data=tv_show_netflix_data)
plt.show()
```



TV-G is having the least rated used to rate the movie both the TV-Y and TV-Y7 are not having any outliers most outliers are having in the rating TV-14 followed by TV-G

In [112]:
```
ax = sns.boxplot(y = 'release_year',data=tv_show_netflix_data)
plt.show()
```



25% is around 2016 50% is around 2018 75% is around 2020 ie., most number of tv shows are produced in the year 2020 and 2021

In [113]:
```
tv_show_netflix_data.groupby('director')['cast'].max().reset_index()
```

Out[113]:

	director	cast
0	Adrien Lagier	Fary
1	Ahmet Katıksız	İlkin Tüfekçi
2	Alan Poul	Tchéky Karyo
3	Alastair Fothergill	David Attenborough
4	Alejandro Lozano	Raúl Méndez
...	...	...
185	Wouter Bouvijn	Zouzou Ben Chikha
186	YC Tom Lee	Wen Chen-ling
187	Yasuhiro Irie	Yuji Ueda
188	Yim Pilsung	Shim Dal-gi
189	Ziad Doueiri	Suzanne Clément

190 rows × 2 columns

The most favorite actor for the director is listed above

In [115]:
```
tv_show_netflix_data.groupby('country')['cast'].max().reset_index()
```

Out[115]:

	country	cast
0	Argentina	Sebastián Wainraich
1	Australia	Tess Haubrich
2	Belgium	Zouzou Ben Chikha
3	Brazil	Yasmin Thayná
4	Canada	Wyatt White
5	China	Zhu Rongrong
6	Colombia	Álvaro Bayona
7	Denmark	William Rützou
8	France	Tony Sampson
9	Germany	Tchéky Karyo
10	Greece	David Attenborough
11	India	Yogi Babu
12	Indonesia	Steve Blum
13	Ireland	Tom Wilkinson
14	Israel	Linor Abargil
15	Italy	Wayne Grayson
16	Japan	Yuzuka Nakaya
17	Malaysia	Wu Kang-jen
18	Mexico	Sebastián Osorio
19	Nigeria	Zynnell Zuh
20	Pakistan	Sanya Shamshad
21	Philippines	Steve Blum
22	Poland	Zofia Wichłacz
23	Russia	Yuri Kuznetsov
24	Saudi Arabia	Zara Albalushi
25	Singapore	Steve Blum
26	South Africa	Linor Abargil
27	South Korea	Yun Seo-hyun
28	Spain	Ángela Molina
29	Taiwan	Yin Chao-te
30	Thailand	Yanin Opassathaworn
31	Turkey	İştar Gökseven
32	United Kingdom	Whoopi Goldberg
33	United States	Zachary Booth

Most number of tv shows acted by the actor in his country

In [116]:
```
tv_show_netflix_data.groupby('listed_in')['director'].max().reset_index()
```

Out[116]:

	listed_in	director
0	Anime Series	Yasuhiro Irie
1	British TV Shows	Toby Haynes
2	Classic & Cult TV	Phil Sgriccia
3	Crime TV Shows	Ziad Doueiri
4	Docuseries	Simon Frederick
5	International TV Shows	Ziad Doueiri
6	Kids' TV	Tony Collingwood
7	Korean TV Shows	Yim Pilsung
8	Reality TV	Michael Simon
9	Romantic TV Shows	Takuya Igarashi
10	Science & Nature TV	Everardo Gout
11	Spanish-Language TV Shows	Mateo Gil
12	Stand-Up Comedy & Talk Shows	Tig Notaro
13	TV Action & Adventure	Thomas Astruc
14	TV Comedies	Tig Notaro
15	TV Dramas	Ziad Doueiri
16	TV Horror	Su I-Hsuan
17	TV Mysteries	Serdar Akar
18	TV Sci-Fi & Fantasy	Toby Haynes
19	TV Shows	Vasanth Sai
20	TV Thrillers	YC Tom Lee
21	Teen TV Shows	Takuya Igarashi

Most number of tv shows Directed by the director in the Genre

In [117]:
```
tv_show_netflix_data.groupby('listed_in')['cast'].max().reset_index()
```

Out[117]:

	listed_in	cast
0	Anime Series	Yusuke Kobayashi
1	British TV Shows	Álvaro Cervantes
2	Classic & Cult TV	Tracy Ann Oberman
3	Crime TV Shows	İlkin Tüfekçi
4	Docuseries	Zachary Booth
5	International TV Shows	İştar Gökseven
6	Kids' TV	Wyatt White
7	Korean TV Shows	Yun Seo-hyun
8	Reality TV	Sue Perkins
9	Romantic TV Shows	İştar Gökseven
10	Science & Nature TV	Sammi Rotibi
11	Spanish-Language TV Shows	Ángela Molina
12	Stand-Up Comedy & Talk Shows	Young-chul Kim
13	TV Action & Adventure	Zhu Rongrong
14	TV Comedies	İştar Gökseven
15	TV Dramas	İlkin Tüfekçi
16	TV Horror	Young Dais
17	TV Mysteries	Özay Fecht
18	TV Sci-Fi & Fantasy	Wentworth Miller
19	TV Shows	Zara Albalushi
20	TV Thrillers	Wen Chen-ling
21	Teen TV Shows	Vathusiri Phuwapunyasiri

Most number of tv shows acted by the actor in the Genre

In [124]:
```
#tv_show_netflix_data.groupby('listed_in')['director'].max().reset_index()
```

In [119]:
```
tv_show_netflix_data.groupby('country')['listed_in'].max().reset_index()
```

Out[119]:

	country	listed_in
0	Argentina	TV Comedies
1	Australia	TV Dramas
2	Belgium	TV Dramas
3	Brazil	TV Mysteries
4	Canada	Teen TV Shows
5	China	TV Dramas
6	Colombia	Spanish-Language TV Shows
7	Denmark	TV Dramas
8	France	TV Dramas
9	Germany	TV Dramas
10	Greece	International TV Shows
11	India	TV Shows
12	Indonesia	TV Horror
13	Ireland	TV Action & Adventure
14	Israel	Docuseries
15	Italy	TV Dramas
16	Japan	Teen TV Shows
17	Malaysia	TV Mysteries
18	Mexico	Spanish-Language TV Shows
19	Nigeria	TV Dramas
20	Pakistan	TV Dramas
21	Philippines	TV Horror
22	Poland	TV Dramas
23	Russia	TV Mysteries
24	Saudi Arabia	TV Shows
25	Singapore	TV Horror
26	South Africa	Docuseries
27	South Korea	TV Mysteries
28	Spain	TV Dramas
29	Taiwan	TV Thrillers
30	Thailand	Teen TV Shows
31	Turkey	TV Thrillers
32	United Kingdom	TV Sci-Fi & Fantasy
33	United States	TV Thrillers

country specific most liked Genre

In [120]:
```
tv_show_netflix_data.groupby('country')['director'].max().reset_index()
```

Out[120]:

	country	director
0	Argentina	Hernán Guerschuny
1	Australia	Mat King
2	Belgium	Wouter Bouvijn
3	Brazil	Pedro Waddington
4	Canada	Tony Collingwood
5	China	He Xiaofeng
6	Colombia	Luis Alberto Restrepo
7	Denmark	Stephen Murray
8	France	Ziad Doueiri
9	Germany	Laïla Marrakchi
10	Greece	Alastair Fothergill
11	India	Vikramaditya Motwane
12	Indonesia	Jay Oliva
13	Ireland	Noam Murro
14	Israel	Cecilia Peck
15	Italy	Stefano Lodovichi
16	Japan	Yasuhiro Irie
17	Malaysia	Quek Shio-chuan
18	Mexico	Manolo Caro
19	Nigeria	Tosin Coker
20	Pakistan	Ehtesham Uddin
21	Philippines	Richard Arellano
22	Poland	Jan Holoubek
23	Russia	Pavel Kostomarov
24	Saudi Arabia	Mohamed al Salman
25	Singapore	Jay Oliva
26	South Africa	Cecilia Peck
27	South Korea	Yim Pilsung
28	Spain	Patricia Font
29	Taiwan	YC Tom Lee
30	Thailand	Sivaroj Kongsakul
31	Turkey	Seren Yüce
32	United Kingdom	Tony Collingwood
33	United States	Vikramaditya Motwane

Most number of Tv shows directed by the director in his country

In [121]:
```
tv_show_netflix_data.groupby('director')['duration'].max().sort_values(ascending=
```

Out[121]:

	director	duration
0	Philippa Lowthorpe	9 Seasons
1	Andy Devonshire	9 Seasons
2	Hayato Date	9 Seasons
3	James Bamford	8 Seasons
4	Jay Chandrasekhar	8 Seasons
...	...	...
185	Kongkiat Khomsiri	1 Season
186	Kobun Shizuno	1 Season
187	Yim Pilsung	1 Season
188	Kim Jong-kwan	1 Season
189	Ziad Doueiri	1 Season

190 rows × 2 columns

most number of seasons produced by the director

This recomendation is for the country Turkey we can draw the similar conclusions for the above data to the remaining countries as well

for Example if we want to recommend a Tv shows in the country Turkey as per the data The triller TV shows are most popular in that country so it is best to release the triller Tv shows in that country.

When comes to the director Seren Yüce is the popular director so the movies directed by him are most likely to get outreacch in Turkey

When comes to the actor İştar Gökseven is the famous actor in the country so it is best to release his movies in Turkey.

In [ ]: