In [887]:

```python
import numpy as np
import pandas as pd
import seaborn as sns

crude_df = pd.read_csv("C:\\Users\\bolla\\OneDrive\\Desktop\\Notes\\netflix.csv")
crude_df
```

Out[887]:

	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

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

8807 rows × 12 columns

In [888]:

```python
print("Size of the data :", crude_df.size , "elements")
```

Size of the data : 105684 elements

In [889]:

```python
### (Problem statement = we have a data of different movies & series along with its directo
#   release , cast , listed_in and their description , some of the valuse in the data are m
#   numerical than catogorical  )
```

In [890]:

```python
crude_df.info()
```

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

In [891]:

```python
crude_df.shape
```

Out[891]:

```
(8807, 12)
```

In [892]:

```python
crude_df.describe()
```

Out[892]:

	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 [893]:

```python
# % of missing values in each column
missing = crude_df.isna().sum()/len(crude_df)*100
missing
```

Out[893]:

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

In [894]:

```python
crude_df = crude_df.drop(['show_id'], axis = 1)
crude_df
```

Out[894]:

	type	title	director	cast	country	date_added	release_year	rating	durati
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 r
1	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	Seasc
2	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	Seas
3	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	Seas
4	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	Seasc
...	...	...	...	...	...	...	...	...	
8802	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 r
8803	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	Seasc
8804	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 r

	type	title	director	cast	country	date_added	release_year	rating	durati
8805	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 r
8806	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 r

8807 rows × 11 columns
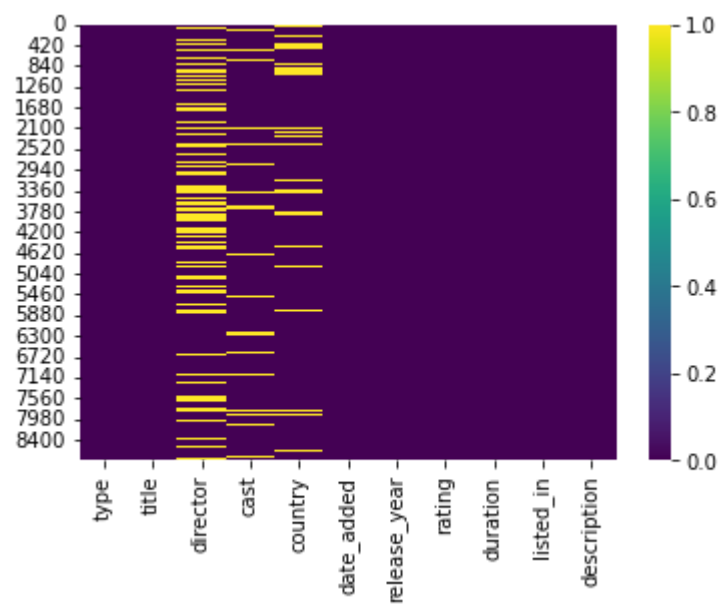

In [895]:

```
sns.heatmap(crude_df.isna(),cmap = 'viridis')
```

Out[895]:

```
<AxesSubplot:>
```


In [896]:

```
tvshows = crude_df.loc[crude_df["type"]=="TV Show"]
tvshows = tvshows.reset_index(drop=True)
print("number of tvshows",len(tvshows))
```

```
number of tvshows 2676
```

In [897]:

```python
movies = crude_df.loc[crude_df["type"]=="Movie"]
movies = movies.reset_index(drop=True)
print("number of movies",len(movies))
```

number of movies 6131

In [898]:

```python
movies.director.nunique()
```

Out[898]:

4354

In [899]:

```python
tvshows.director.nunique()
```

Out[899]:

222

In [900]:

```python
movies.director.isna().sum()
```

Out[900]:

188

In [901]:

```python
tvshows.director.isna().sum()
```

Out[901]:

2446

In [902]:

```
movies['director'].fillna(movies['director'].mode()[0], inplace = True)
movies
```

Out[902]:

	type	title	director	cast	country	date_added	release_year	ratin
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG 1
1	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	NaN	September 24, 2021	2021	P
2	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV M
3	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG 1
4	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	2021	TV M
...	...	...	...	...	...	...	...	
6126	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV M
6127	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	
6128	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	
6129	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	P

	type	title	director	cast	country	date_added	release_year	ratin
6130	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-1

6131 rows × 11 columns

In [903]:

```
tvshows['director'].fillna(tvshows['director'].mode()[0], inplace = True)
tvshows
```

Out[903]:

	type	title	director	cast	country	date_added	release_year	rating	dura
0	TV Show	Blood & Water	Alastair Fothergill	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	Sea:
1	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	Sea
2	TV Show	Jailbirds New Orleans	Alastair Fothergill	NaN	NaN	September 24, 2021	2021	TV-MA	Sea
3	TV Show	Kota Factory	Alastair Fothergill	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	Sea:
4	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	Sea
...	...	...	...	...	...	...	...	...	
2671	TV Show	Yu-Gi-Oh! Arc-V	Alastair Fothergill	Mike Liscio, Emily Bauer, Billy Bob Thompson, ...	Japan, Canada	May 1, 2018	2015	TV-Y7	Sea:
2672	TV Show	Yunus Emre	Alastair Fothergill	Gökhan Atalay, Payidar Tüfekçioglu, Baran Akbu...	Turkey	January 17, 2017	2016	TV-PG	Sea
2673	TV Show	Zak Storm	Alastair Fothergill	Michael Johnston, Jessica Gee-George, Christin...	United States, France, South Korea, Indonesia	September 13, 2018	2016	TV-Y7	Sea:

	type	title	director	cast	country	date_added	release_year	rating	dura
2674	TV Show	Zindagi Gulzar Hai	Alastair Fothergill	Sanam Saeed, Fawad Khan, Ayesha Omer, Mehreen ...	Pakistan	December 15, 2016	2012	TV-PG	Sea
2675	TV Show	Zombie Dumb	Alastair Fothergill	NaN	NaN	July 1, 2019	2018	TV-Y7	Sea

2676 rows × 11 columns

In [904]:

```python
movies['cast'].fillna(movies['cast'].mode()[0], inplace = True)
tvshows['cast'].fillna(tvshows['cast'].mode()[0], inplace = True)
```

In [905]:

```python
movies['duration'] = movies['duration'].apply((lambda x : str(x).split(" ")))
movies['duration'] = movies['duration'].apply(lambda x : x[0])
#movies
```

In [906]:

```python
tvshows['duration'] = tvshows['duration'].apply((lambda x : str(x).split(" ")))
tvshows['duration'] = tvshows['duration'].apply(lambda x : x[0])

#tvshows
```

In [907]:

```python
df = pd.concat([movies,tvshows])
df
```

Out[907]:

	type	title	director	cast	country	date_added	release_year	rating
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	United States	September 25, 2021	2020	PG-13
1	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	NaN	September 24, 2021	2021	PG
2	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA
3	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG-13
4	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	2021	TV-MA
...	...	...	...	...	...	...	...	...
2671	TV Show	Yu-Gi-Oh! Arc-V	Alastair Fothergill	Mike Liscio, Emily Bauer, Billy Bob Thompson, ...	Japan, Canada	May 1, 2018	2015	TV-Y7
2672	TV Show	Yunus Emre	Alastair Fothergill	Gökhan Atalay, Payidar Tüfekçioglu, Baran Akbu...	Turkey	January 17, 2017	2016	TV-PG
2673	TV Show	Zak Storm	Alastair Fothergill	Michael Johnston, Jessica Gee-George, Christin...	United States, France, South Korea, Indonesia	September 13, 2018	2016	TV-Y7
2674	TV Show	Zindagi Gulzar Hai	Alastair Fothergill	Sanam Saeed, Fawad Khan, Ayesha Omer, Mehreen ...	Pakistan	December 15, 2016	2012	TV-PG

	type	title	director	cast	country	date_added	release_year	rating
2675	TV Show	Zombie Dumb	Alastair Fothergill	David Attenborough	NaN	July 1, 2019	2018	TV-Y7

8807 rows × 11 columns

In [908]:

```python
df.isna().sum()/len(df)*100
```

Out[908]:

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

In [909]:

```python
df.country.fillna(df["country"].mode()[0],inplace=True)
```

In [910]:

```
df.dropna(axis =0, how="any", inplace = True)
df
```

Out[910]:

	type	title	director	cast	country	date_added	release_year	rating
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	United States	September 25, 2021	2020	PG-13
1	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	United States	September 24, 2021	2021	PG
2	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA
3	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG-13
4	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	2021	TV-MA
...	...	...	...	...	...	...	...	...
2671	TV Show	Yu-Gi-Oh! Arc-V	Alastair Fothergill	Mike Liscio, Emily Bauer, Billy Bob Thompson, ...	Japan, Canada	May 1, 2018	2015	TV-Y7
2672	TV Show	Yunus Emre	Alastair Fothergill	Gökhan Atalay, Payidar Tüfekçioglu, Baran Akbu...	Turkey	January 17, 2017	2016	TV-PG
2673	TV Show	Zak Storm	Alastair Fothergill	Michael Johnston, Jessica Gee-George, Christin...	United States, France, South Korea, Indonesia	September 13, 2018	2016	TV-Y7
2674	TV Show	Zindagi Gulzar Hai	Alastair Fothergill	Sanam Saeed, Fawad Khan, Ayesha Omer, Mehreen ...	Pakistan	December 15, 2016	2012	TV-PG

	type	title	director	cast	country	date_added	release_year	rating
2675	TV Show	Zombie Dumb	Alastair Fothergill	David Attenborough	United States	July 1, 2019	2018	TV-Y7

8793 rows × 11 columns

In [911]:

```python
df.isna().sum()/len(df)*100
```

Out[911]:

```
type            0.0
title           0.0
director        0.0
cast            0.0
country         0.0
date_added      0.0
release_year    0.0
rating          0.0
duration        0.0
listed_in       0.0
description     0.0
dtype: float64
```

In [912]:

```python
df["date_added"] = pd.to_datetime(df["date_added"])
df
```

Out[912]:

	type	title	director	cast	country	date_added	release_year	rating	duration	
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	United States	2021-09-25	2020	PG-13	90	Do
1	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	United States	2021-09-24	2021	PG	91	Fa
2	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano	United States, Ghana, Burkina	2021-09-24	1993	TV-MA	125	

In [913]:

```python
df['duration'] = df['duration'].apply(lambda x : float(x))

print("average no.of seasons tvshow : " ,df[df["type"]== "TV Show"].duration.mean())
print("average mins of movie : " ,df[df["type"]== "Movie"].duration.mean())
```

```
average no.of seasons tvshow :  1.7518768768768769
average mins of movie :  99.58488410055502
```
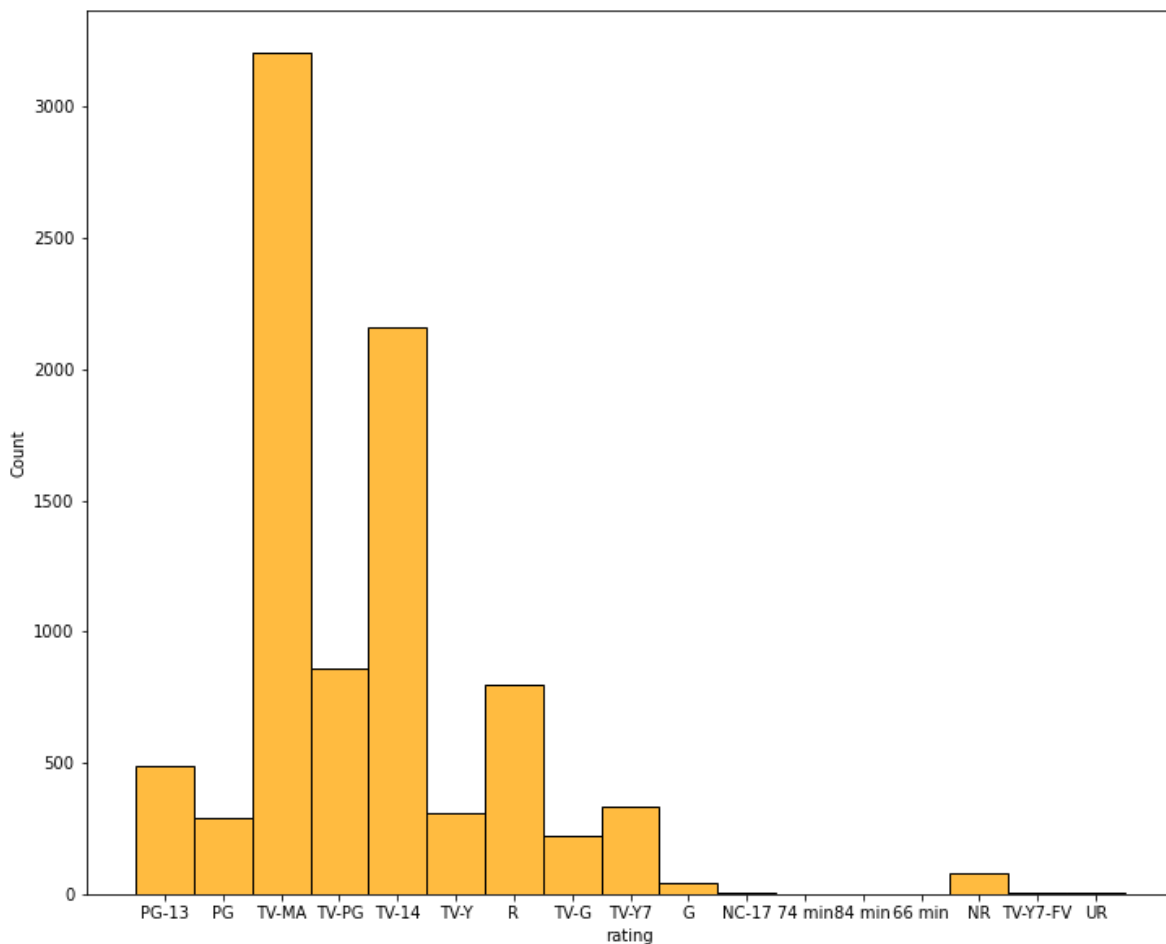
In [914]:

```python
import matplotlib.pyplot as plt

plt.figure(figsize=(12,10))
sns.histplot(df['rating'], color ="orange")
```
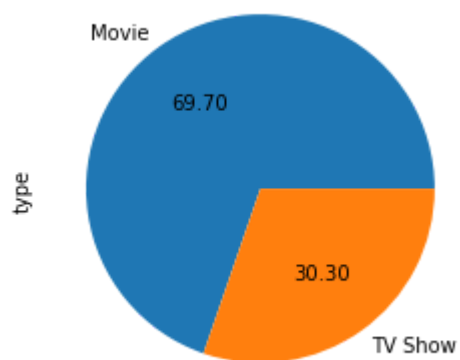
Out[914]:

```
<AxesSubplot:xlabel='rating', ylabel='Count'>
```

In [915]:

```python
df['type'].value_counts().plot(kind='pie',autopct="%.2f")
```

Out[915]:

```
<AxesSubplot:ylabel='type'>
```

In [916]:

```python
plt.figure(figsize=(12,10))
sns.countplot(data = df , x = "rating" , hue = "type")
```
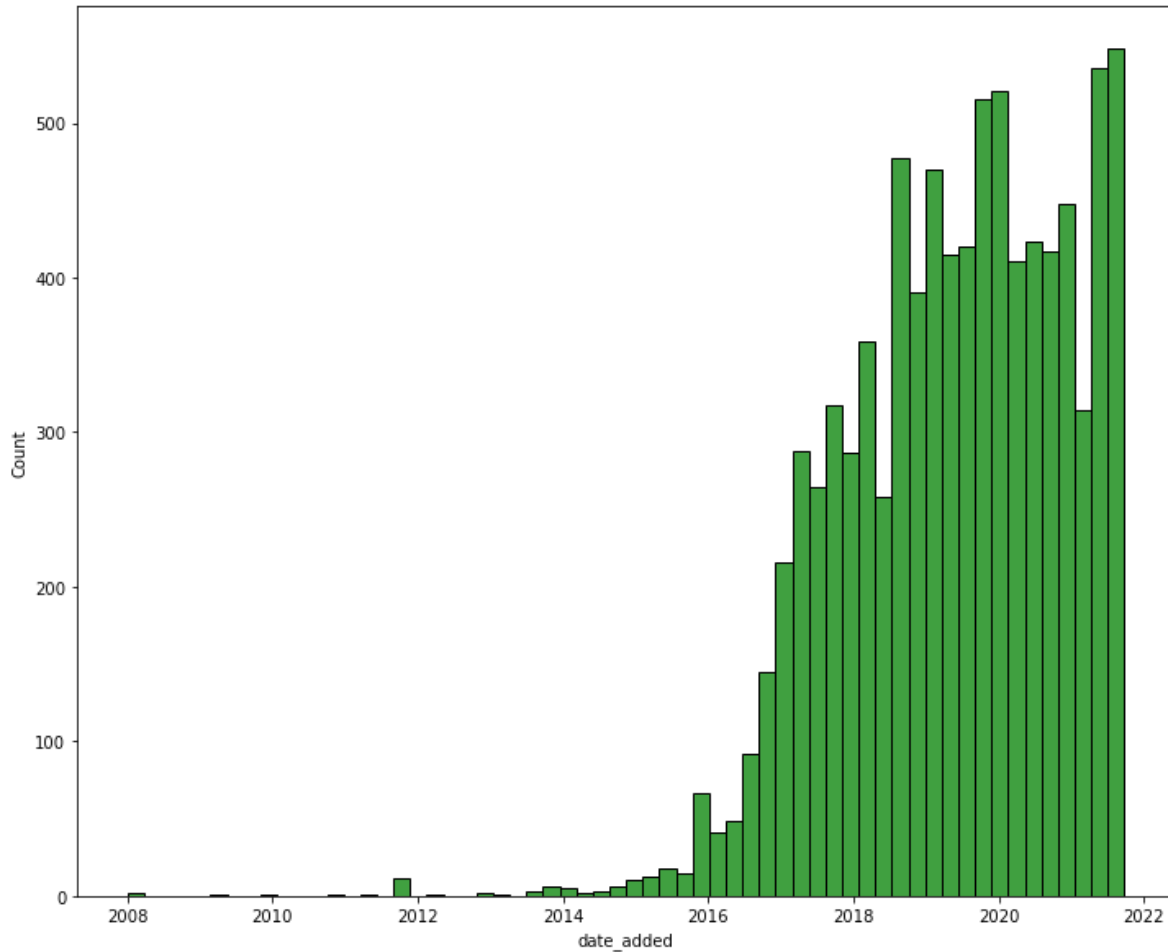
Out[916]:

```
<AxesSubplot:xlabel='rating', ylabel='count'>
```
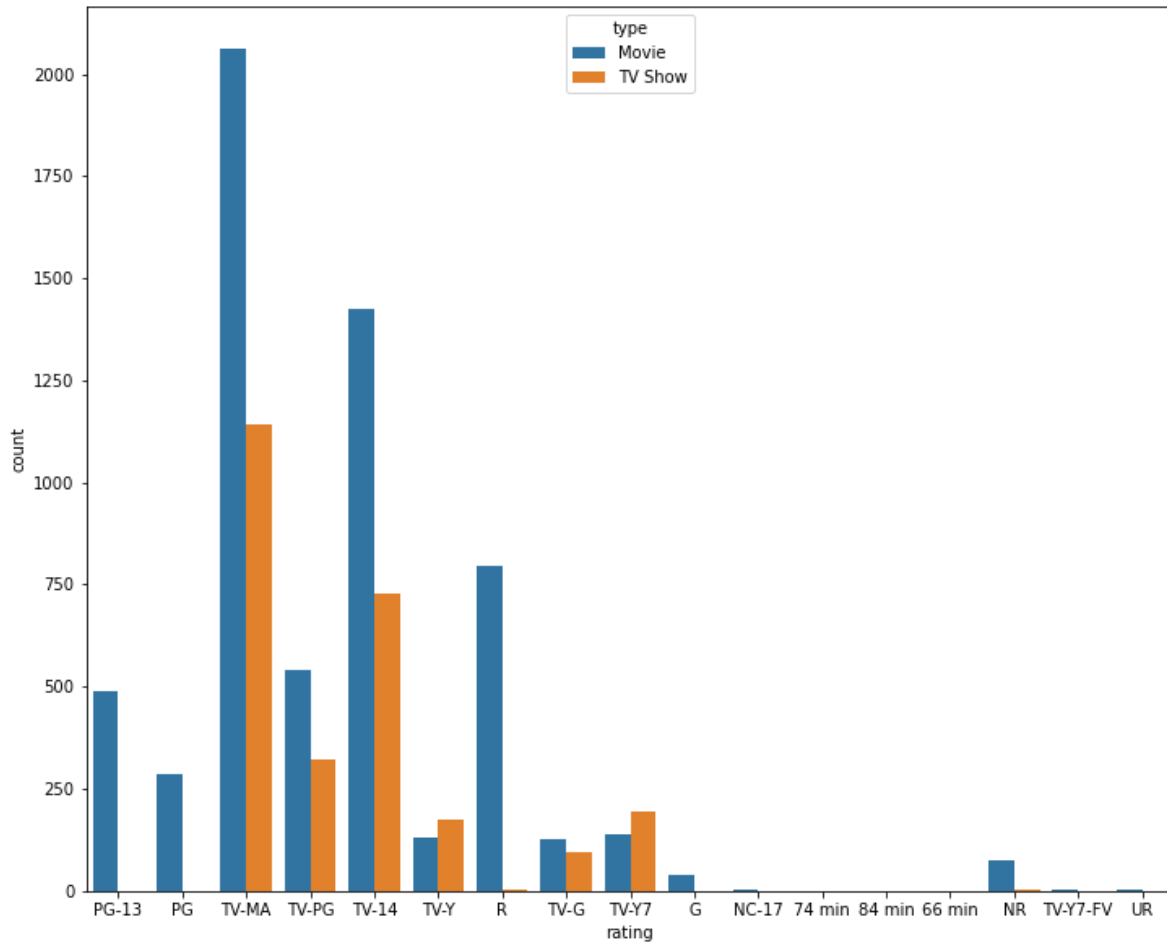
In [917]:

```python
plt.figure(figsize=(12,10))
sns.histplot(df['date_added'],color = "green")
```

Out[917]:

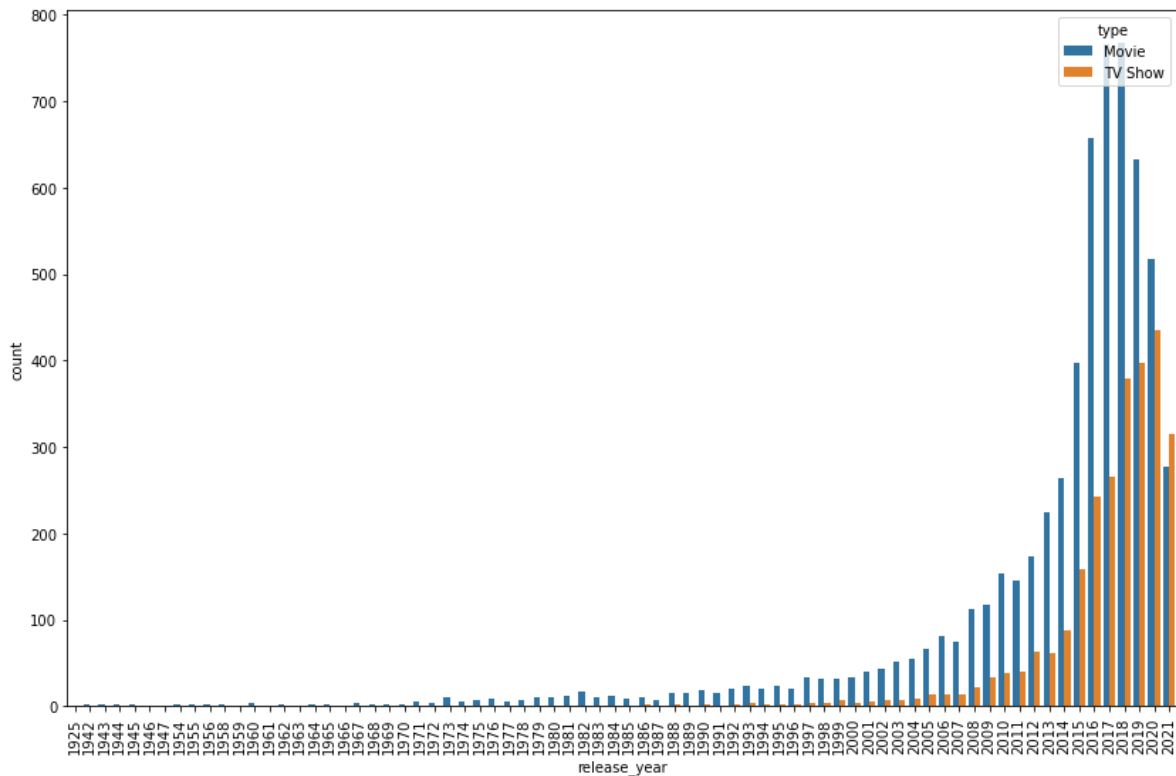```
<AxesSubplot:xlabel='date_added', ylabel='Count'>
```

In [918]:

```python
plt.figure(figsize=(14,9))
plt.xticks(rotation=90)
sns.countplot(data = df , x = "release_year" , hue = "type")
```

Out[918]:

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

In [919]:

```python
plt.figure(figsize=(10,8))
plt.xticks(rotation=90)
sns.scatterplot(data = df , x = "release_year", y = "date_added", hue = "type")
```

Out[919]:

```
<AxesSubplot:xlabel='release_year', ylabel='date_added'>
```
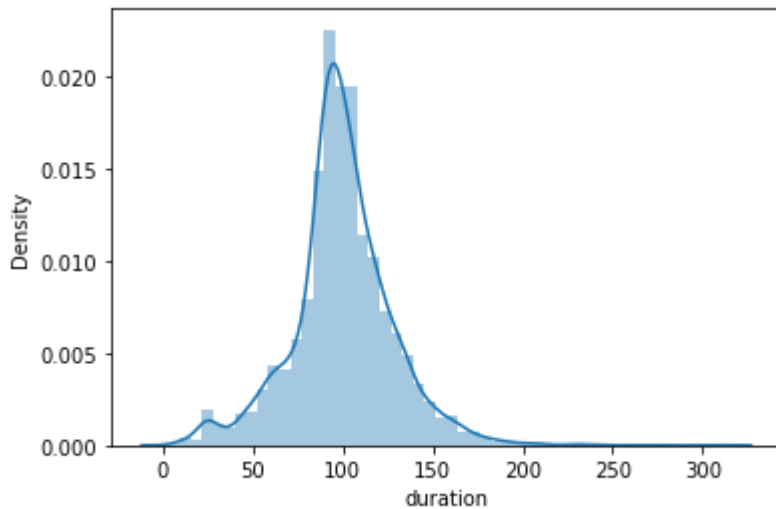
In [920]:

```python
sns.distplot(movies["duration"])
```

C:\Users\bolla\anaconda3\lib\site-packages\seaborn\distributions.py:2619: Fu
tureWarning: `distplot` is a deprecated function and will be removed in a fu
ture 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)

Out[920]:

```
<AxesSubplot:xlabel='duration', ylabel='Density'>
```
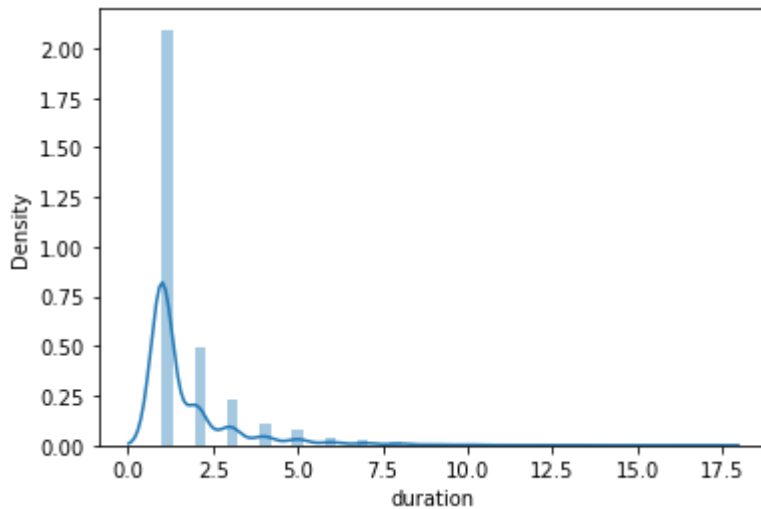
In [921]:

```
sns.distplot(tvshows["duration"])
```

C:\Users\bolla\anaconda3\lib\site-packages\seaborn\distributions.py:2619: Fu
tureWarning: `distplot` is a deprecated function and will be removed in a fu
ture 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)

Out[921]:

```
<AxesSubplot:xlabel='duration', ylabel='Density'>
```

In [922]:

```python
df.head(10)
```

Out[922]:

	type	title	director	cast	country	date_added	release_year	rating	dur
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	United States	2021-09-25	2020	PG-13	
1	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	United States	2021-09-24	2021	PG	
2	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	2021-09-24	1993	TV-MA	
3	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	2021-09-24	2021	PG-13	
4	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	2021-09-23	2021	TV-MA	
5	Movie	Confessions of an Invisible Girl	Bruno Garotti	Klara Castanho, Lucca Picon, Júlia Gomes, Marc...	United States	2021-09-22	2021	TV-PG	
6	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	Pedro de Echave García, Pablo Azorín Williams	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	United States	2021-09-22	2020	TV-MA	
7	Movie	Intrusion	Adam Salky	Freida Pinto, Logan Marshall-Green, Robert Joh...	United States	2021-09-22	2021	TV-14	
8	Movie	Avvai Shanmughi	K.S. Ravikumar	Kamal Hassan, Meena, Gemini Ganesan, Heera Raj...	United States	2021-09-21	1996	TV-PG	

	type	title	director	cast	country	date_added	release_year	rating	dur
9	Movie	Go! Go! Cory Carson: Chrissy Takes the Wheel	Alex Woo, Stanley Moore	Maisie Benson, Paul Killam, Kerry Gudjohnsen, ...	United States	2021-09-21	2021	TV-Y	

In [923]:

```python
df["listed_in"].nunique()
```

Out[923]:

513

In [924]:

```python
dup = df[df.duplicated(keep = "first")]
dup
# no duplicate rows found
```

Out[924]:

	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	desc

In [925]:

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8793 entries, 0 to 2675
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   type          8793 non-null   object
 1   title         8793 non-null   object
 2   director      8793 non-null   object
 3   cast          8793 non-null   object
 4   country       8793 non-null   object
 5   date_added    8793 non-null   datetime64[ns]
 6   release_year  8793 non-null   int64
 7   rating        8793 non-null   object
 8   duration      8790 non-null   float64
 9   listed_in     8793 non-null   object
 10  description   8793 non-null   object
dtypes: datetime64[ns](1), float64(1), int64(1), object(8)
memory usage: 1.1+ MB
```

In [926]:

```python
# top 20 directors with highest number of movies/tvshows
df.groupby("director")["title"].count().sort_values(ascending=False).head(20)
```

Out[926]:

```
director
Alastair Fothergill       2437
Rajiv Chilaka              206
Raúl Campos, Jan Suter      18
Suhas Kadav                 16
Marcus Raboy                16
Jay Karas                   14
Cathy Garcia-Molina         13
Jay Chapman                 12
Youssef Chahine             12
Martin Scorsese             12
Steven Spielberg            11
Don Michael Paul            10
David Dhawan                 9
Johnnie To                   8
Hakan Algül                  8
Kunle Afolayan               8
Lance Bangs                  8
Shannon Hartman              8
Troy Miller                  8
Fernando Ayllón              8
Name: title, dtype: int64
```

In [927]:

```python
#top 10 countries with highest number of movies/tvshows
df.groupby("country")["title"].count().sort_values(ascending=False).head(10)
```

Out[927]:

```
country
United States     3641
India              972
United Kingdom     418
Japan              243
South Korea        199
Canada             181
Spain              145
France             124
Mexico             110
Egypt              106
Name: title, dtype: int64
```

In [928]:

```
ind = df.loc[df["country"]=="India"]
ind = ind.reset_index(drop=True)
ind
```

Out[928]:

	type	title	director	cast	country	date_added	release_year	rating	c
0	Movie	Jeans	S. Shankar	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...	India	2021-09-21	1998	TV-14	
1	Movie	Angamaly Diaries	Lijo Jose Pellissery	Antony Varghese, Reshma Rajan, Binny Rinky Ben...	India	2021-09-05	2017	TV-14	
2	Movie	Anjaam	Rahul Rawail	Madhuri Dixit, Shah Rukh Khan, Tinnu Anand, Jo...	India	2021-09-02	1994	TV-14	
3	Movie	Dhanak	Nagesh Kukunoor	Krrish Chhabria, Hetal Gada, Vipin Sharma, Gul...	India	2021-09-02	2015	TV-PG	
4	Movie	Gurgaon	Shanker Raman	Akshay Oberoi, Pankaj Tripathi, Ragini Khanna,...	India	2021-09-02	2017	TV-14	
...	...	...	...	...	...	...	...	...	
967	TV Show	Thackeray	Alastair Fothergill	David Attenborough	India	2019-05-25	2019	TV-MA	
968	TV Show	The Calling	Bumpy	David Attenborough	India	2019-04-01	2018	TV-MA	
969	TV Show	The Golden Years with Javed Akhtar	Alastair Fothergill	David Attenborough	India	2017-06-01	2016	TV-G	
970	TV Show	The House That Made Me	Alastair Fothergill	Vinay Pathak	India	2017-03-31	2015	TV-PG	

	type	title	director	cast	country	date_added	release_year	rating	c
971	TV Show	Yeh Meri Family	Alastair Fothergill	Vishesh Bansal, Mona Singh, Akarsh Khurana, Ah...	India	2018-08-31	2018	TV-PG	

972 rows × 11 columns

In [929]:

```
ind["type"].value_counts() , ## in India , movies have more domination when compared to tv
```
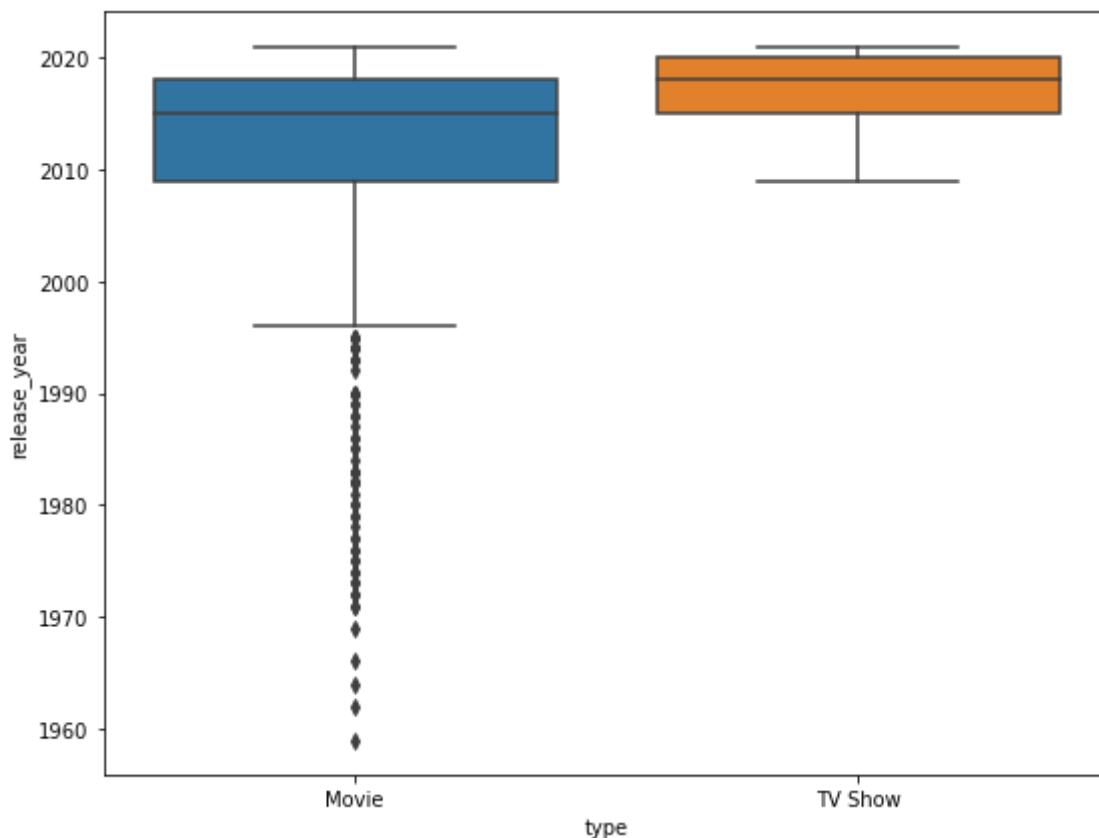
Out[929]:

```
(Movie      893
 TV Show     79
 Name: type, dtype: int64,)
```

In [930]:

```
plt.figure(figsize=(9,7))

sns.boxplot(data=ind, x = "type", y = "release_year")

#there are outliers in movies in Indian data
```

Out[930]:

```
<AxesSubplot:xlabel='type', ylabel='release_year'>
```

In [931]:

```python
dr=tvshows.groupby("director").agg({"title" : "nunique"}).reset_index().sort_values(by=["ti
```

In [932]:

```python
dr.head(10)
```

Out[932]:

	director	title
5	Alastair Fothergill	2449
194	Stan Lathan	2
74	Iginio Straffi	2
177	Rob Seidenglanz	2
108	Ken Burns	2
187	Shin Won-ho	2
72	Hsu Fu-chun	2
144	Miguel Conde	1
145	Mike Flanagan	1
151	Norm Hiscock, Gary Howsam, Mike Smith, John Pa...	1

In [933]:

```python
ind["cast"] = ind["cast"].apply(lambda x : str(x).split(','))
```

In [934]:

```python
ind1=ind.explode("cast")
```

In [935]:

```python
ind1[["director","cast"]].mode()
# highest number of combo of a director and cast in indian country
```

Out[935]:

	director	cast
0	Alastair Fothergill	Anupam Kher

In [936]:

```python
plt.figure(figsize=(12,10))


sns.pairplot(data=ind, hue = "type")
```
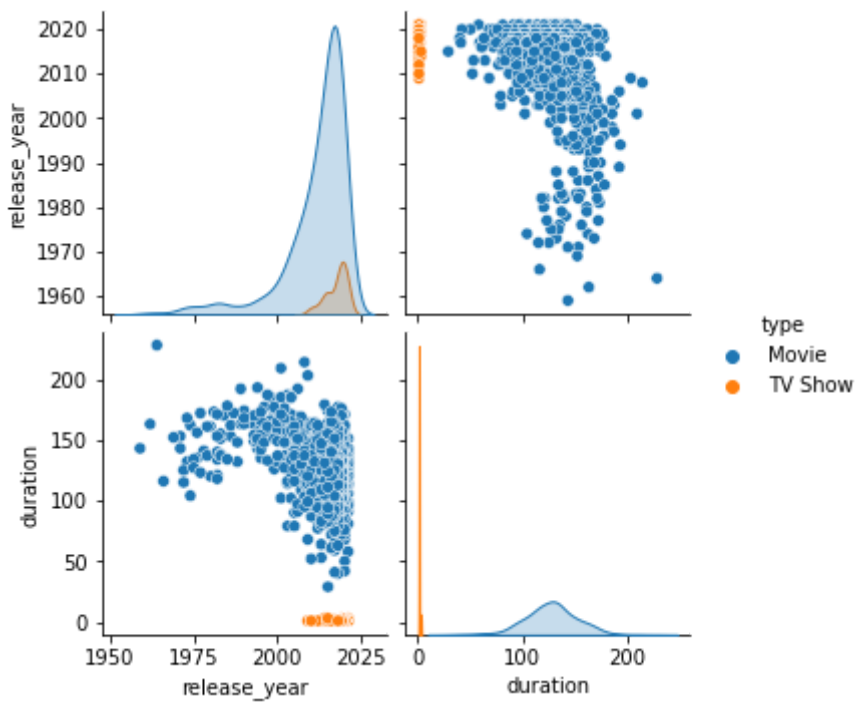
Out[936]:

```
<seaborn.axisgrid.PairGrid at 0x2381108d220>
```

```
<Figure size 864x720 with 0 Axes>
```

In [937]:

```python
df["listed_in"] = df["listed_in"].apply(lambda x : str(x).split(','))
df
```

Out[937]:

	type	title	director	cast	country	date_added	release_year	rating
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	United States	2021-09-25	2020	PG-13
1	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	United States	2021-09-24	2021	PG
2	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	2021-09-24	1993	TV-MA
3	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	2021-09-24	2021	PG-13
4	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	2021-09-23	2021	TV-MA
...	...	...	...	...	...	...	...	...
2671	TV Show	Yu-Gi-Oh! Arc-V	Alastair Fothergill	Mike Liscio, Emily Bauer, Billy Bob Thompson, ...	Japan, Canada	2018-05-01	2015	TV-Y7
2672	TV Show	Yunus Emre	Alastair Fothergill	Gökhan Atalay, Payidar Tüfekçioglu, Baran Akbu...	Turkey	2017-01-17	2016	TV-PG
2673	TV Show	Zak Storm	Alastair Fothergill	Michael Johnston, Jessica Gee-George, Christin...	United States, France, South Korea, Indonesia	2018-09-13	2016	TV-Y7
2674	TV Show	Zindagi Gulzar Hai	Alastair Fothergill	Sanam Saeed, Fawad Khan, Ayesha Omer, Mehreen ...	Pakistan	2016-12-15	2012	TV-PG

	type	title	director	cast	country	date_added	release_year	rating
2675	TV Show	Zombie Dumb	Alastair Fothergill	David Attenborough	United States	2019-07-01	2018	TV-Y7

8793 rows × 11 columns

◄ ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ ▶

In [938]:

```python
df1=df.explode("listed_in").reset_index(drop=True)
```

In [939]:

```python
df1.head()
```

Out[939]:

	type	title	director	cast	country	date_added	release_year	rating	duration
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	United States	2021-09-25	2020	PG-13	90.0
1	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	United States	2021-09-24	2021	PG	91.0
2	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	2021-09-24	1993	TV-MA	125.0
3	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	2021-09-24	1993	TV-MA	125.0
4	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	2021-09-24	1993	TV-MA	125.0

In [940]:

```python
df1["listed_in"].value_counts()
```

Out[940]:

```
 International Movies       2624
Dramas                     1599
Comedies                   1210
Action & Adventure          859
Documentaries               829
                           ...
Romantic Movies               3
Spanish-Language TV Shows     2
Sports Movies                 1
LGBTQ Movies                  1
TV Sci-Fi & Fantasy           1
Name: listed_in, Length: 73, dtype: int64
```

In [941]:

```python
len(df1[df1["listed_in"].isin(["Comedies","TV Comedies"])])
# total number of movies & tvshows in comedy genre
```

Out[941]:

1329

In [942]:

```python
df1.rating.unique()
```
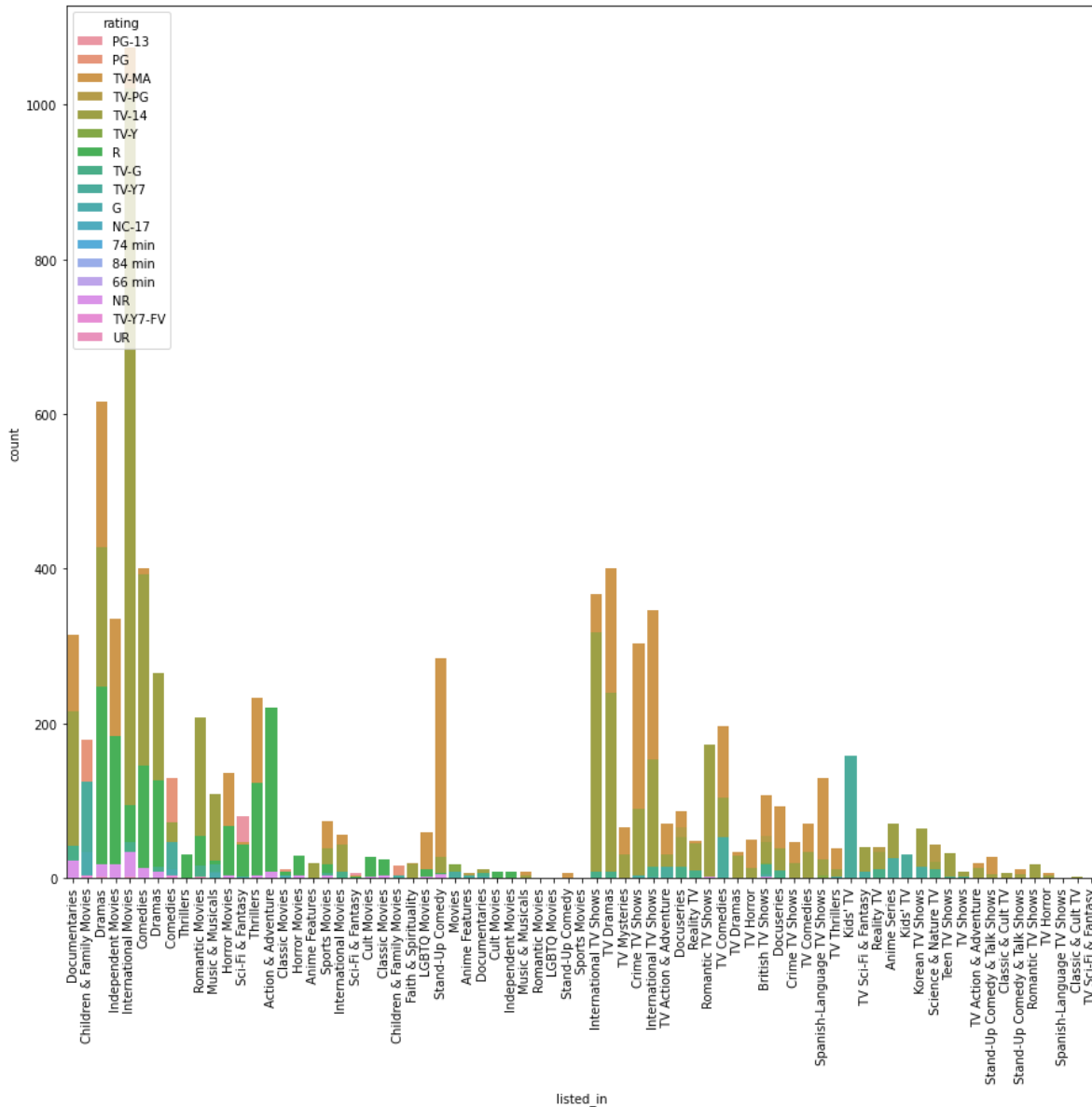
Out[942]:

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

In [943]:

```
plt.figure(figsize=(15,13))
plt.xticks(rotation=90)
sns.countplot(data=df1,x="listed_in",hue="rating",dodge=False)
# plot showing the relation betwen genre and rating of movie
```

Out[943]:

```
<AxesSubplot:xlabel='listed_in', ylabel='count'>
```

In [944]:

```
df1.rating.value_counts()
```

Out[944]:

```
TV-MA        7306
TV-14        5239
TV-PG        1905
R            1582
PG-13         998
PG            607
TV-Y7         527
TV-G          459
TV-Y          405
NR            165
G              73
TV-Y7-FV       13
UR              8
NC-17           7
74 min          1
84 min          1
66 min          1
Name: rating, dtype: int64
```

In [945]:

```python
#converting rating into numerical for corelating
df1["rating"] = pd.factorize(df1["rating"])[0]
df1
```

Out[945]:

	type	title	director	cast	country	date_added	release_year	rating
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	United States	2021-09-25	2020	0
1	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	United States	2021-09-24	2021	1
2	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	2021-09-24	1993	2
3	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	2021-09-24	1993	2
4	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	2021-09-24	1993	2
...	...	...	...	...	...	...	...	...
19292	TV Show	Zindagi Gulzar Hai	Alastair Fothergill	Sanam Saeed, Fawad Khan, Ayesha Omer, Mehreen ...	Pakistan	2016-12-15	2012	3
19293	TV Show	Zindagi Gulzar Hai	Alastair Fothergill	Sanam Saeed, Fawad Khan, Ayesha Omer, Mehreen ...	Pakistan	2016-12-15	2012	3
19294	TV Show	Zombie Dumb	Alastair Fothergill	David Attenborough	United States	2019-07-01	2018	8

	type	title	director	cast	country	date_added	release_year	rating
19295	TV Show	Zombie Dumb	Alastair Fothergill	David Attenborough	United States	2019-07-01	2018	8
19296	TV Show	Zombie Dumb	Alastair Fothergill	David Attenborough	United States	2019-07-01	2018	8

19297 rows × 11 columns

In [946]:

```
df1.rating.value_counts()
```

Out[946]:

```
2     7306
4     5239
3     1905
6     1582
0      998
1      607
8      527
7      459
5      405
14     165
9       73
15      13
16       8
10       7
11       1
12       1
13       1
Name: rating, dtype: int64
```

In [947]:

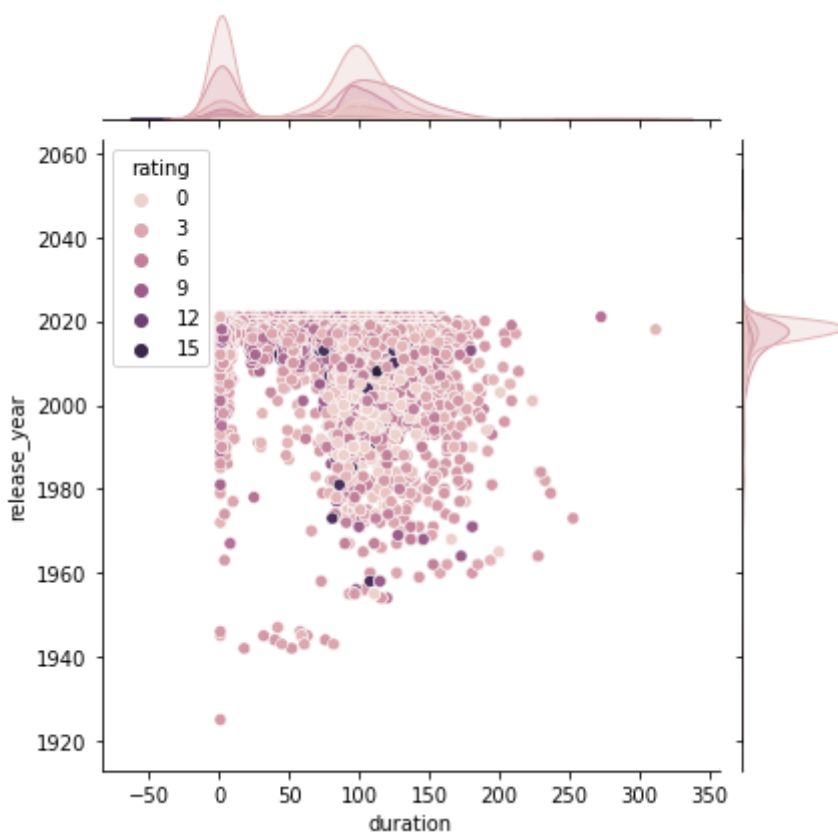```
#TV-MA = 2
#TV-14 = 4
#TV-PG = 3
#R   = 6
#PG-13 = 0
#PG   = 1
#TV-Y7 = 8
#TV-G   = 7
#TV-Y   = 5
#NR   = 14
#G   = 9
#TV-Y7-FV = 15
#UR   = 16
#NC-17   = 10
#74 min   = 11
#84 min   = 12
#66 min   = 13
```

In [948]:

```
sns.jointplot(data=df1,x="duration",y="release_year",hue="rating")
```

Out[948]:

```
<seaborn.axisgrid.JointGrid at 0x238129990a0>
```

In [949]:

```python
df.release_year.value_counts()
```

Out[949]:

```
2018    1146
2017    1031
2019    1030
2020     953
2016     901
        ...
1947       1
1961       1
1966       1
1959       1
1925       1
Name: release_year, Length: 74, dtype: int64
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: