

In [1]:

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

In [2]:

```
df=pd.read_csv('netflix_titles.csv')
```

In [3]:

```
df.head()
```

Out[3]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	d
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	5
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	5

In [4]:

```
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 [5]:

```
df.shape
```

Out[5]:

```
(8807, 12)
```

In [6]:

```
df.describe()
```

Out[6]:

	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 [7]:

```
df.describe(include='object').T
```

Out[7]:

	count	unique	top	freq
show_id	8807	8807	s1	1
type	8807	2	Movie	6131
title	8807	8807	Dick Johnson Is Dead	1
director	6173	4528	Rajiv Chilaka	19
cast	7982	7692	David Attenborough	19
country	7976	748	United States	2818
date_added	8797	1767	January 1, 2020	109
rating	8803	17	TV-MA	3207
duration	8804	220	1 Season	1793
listed_in	8807	514	Dramas, International Movies	362
description	8807	8775	Paranormal activity at a lush, abandoned prope...	4

In [8]:

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

Out[8]:

```
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 [9]:

```
df['type'].value_counts(normalize=True)*100
```

Out[9]:

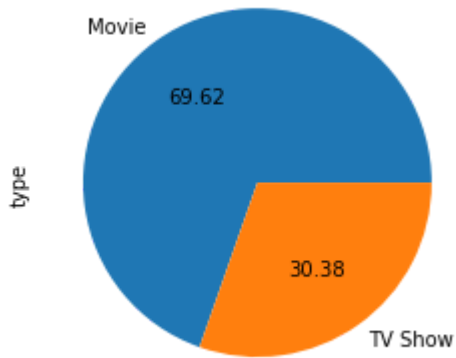
```
Movie      69.615079
TV Show    30.384921
Name: type, dtype: float64
```

In [10]:

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

Out[10]:

<AxesSubplot:ylabel='type'>



In [11]:

df.head()

Out[11]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	d
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	5
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	5

In [12]:

constraint=df['cast'].apply(lambda x: str(x).split(',')).tolist()

In [13]:

df_new=pd.DataFrame(constraint,index=df['title'])

In [14]:

df_new=df_new.stack()

In [15]:

```
df_new=pd.DataFrame(df_new)
```

In [16]:

```
df_new
```

Out[16]:

0

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

64951 rows × 1 columns

In [17]:

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

In [18]:

```
df_new=df_new[['title',0]]
```

In [19]:

```
df_new.columns=['title','cast']
```

In [20]:

```
df_new
```

Out[20]:

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

64951 rows × 2 columns

In [21]:

```
df_cast=df.merge(df_new, on='title', how='inner')
```

In [22]:

```
df_cast[df_cast['cast_y']=='Anupam Kher']
```

Out[22]:

	show_id	type	title	director	cast_x	country	date_added	release_year	re
20434	s2565	Movie	One Day: Justice Delivered	Ashok Nanda	Anupam Kher, Esha Gupta, Kumud Mishra, Kashyap...	India	May 3, 2020	2019	T
44961	s6064	Movie	A Wednesday	Neeraj Pandey	Anupam Kher, Naseeruddin Shah, Jimmy Shergill,...	India	November 1, 2018	2008	
53074	s7187	Movie	Khosla Ka Ghosla	Dibakar Banerjee	Anupam Kher, Boman Irani, Parvin Dabas, Tara S...	India	August 2, 2018	2006	T
64642	s8770	Movie	Y.M.I.: Yeh Mera India	N. Chandra	Anupam Kher, Atul Kulkarni, Aakash Pandey, Mil...	India	March 1, 2018	2008	

In [23]:

```
x=pd.DataFrame(df_cast.groupby('cast_y')['title'].count())
```

In [24]:

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

In [25]:

```
x[x['cast_y']=='Anupam Kher']
```

Out[25]:

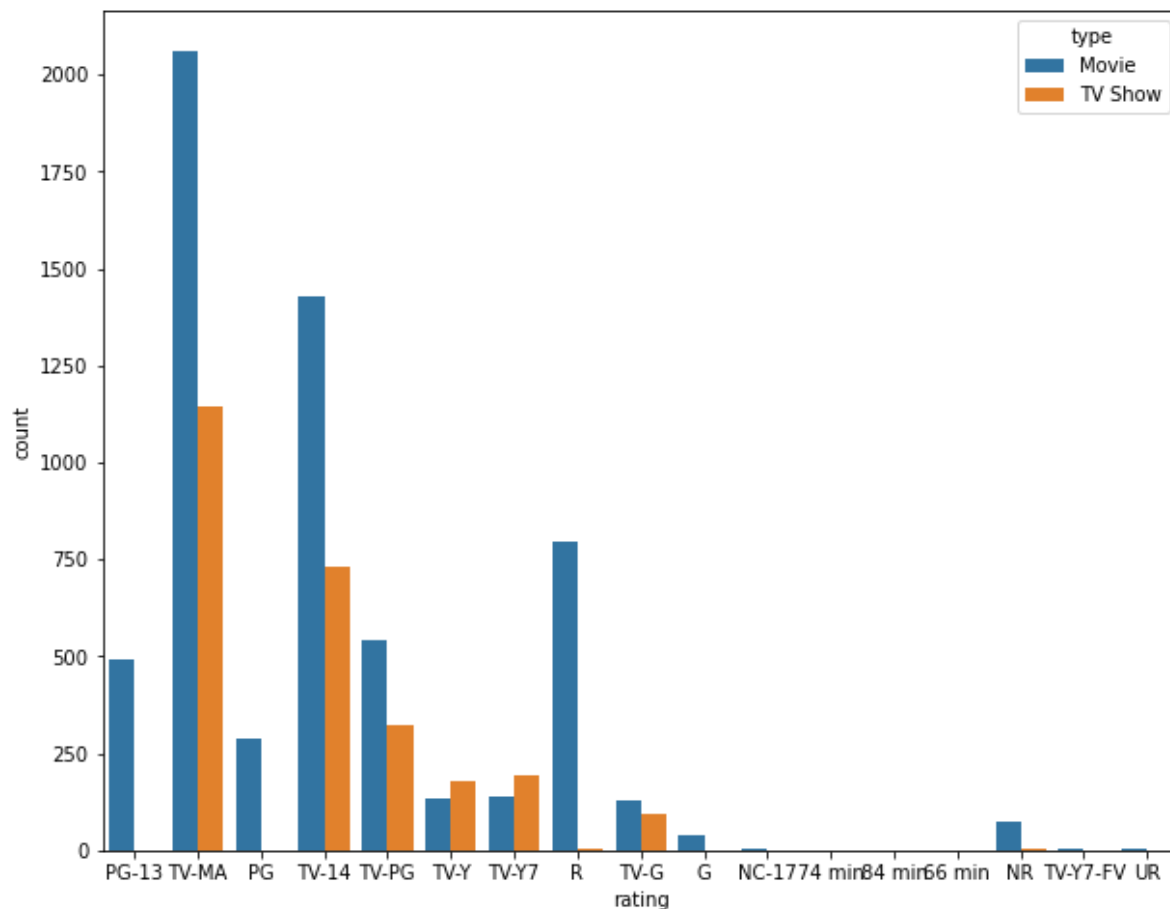
	cast_y	title
34330	Anupam Kher	4

In [26]:

```
import seaborn as sns
import matplotlib.pyplot as plt
plt.figure(figsize=(10,8))
sns.countplot(x='rating', hue='type', data=df)
```

Out[26]:

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

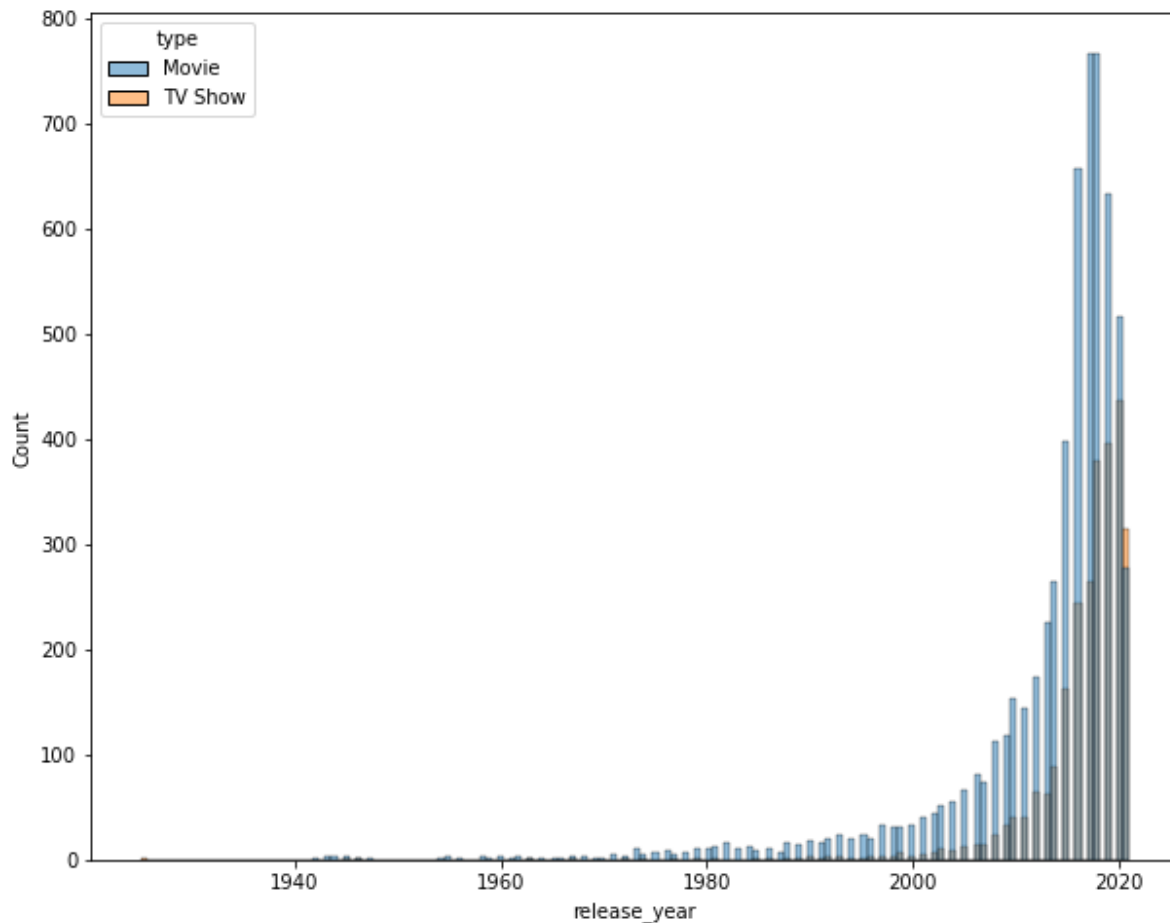


In [27]:

```
plt.figure(figsize=(10,8))  
sns.histplot(x='release_year',hue='type',data=df)
```

Out[27]:

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



In [28]:

```
df['date_add']=pd.to_datetime(df['date_added'], errors='coerce')
```

In [29]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 13 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 
12   date_add        8797 non-null   datetime64[ns]
dtypes: datetime64[ns](1), int64(1), object(11)
memory usage: 894.6+ KB
```

In [30]:

```
df.head()
```

Out[30]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	d
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	5
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	5

In [31]:

```
df['month']=df['date_add'].dt.month.fillna(-1)
```

In [32]:

```
df['month']=df['month'].astype('int64')
```

In [33]:

```
df.head()
```

Out[33]:

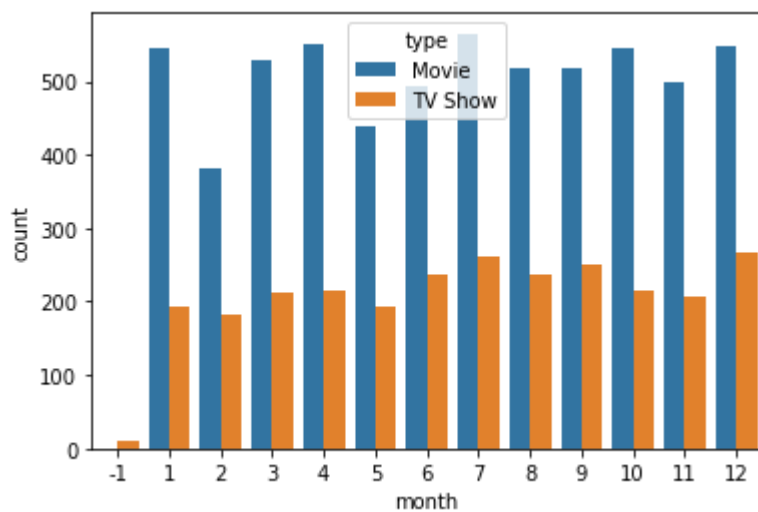
	show_id	type	title	director	cast	country	date_added	release_year	rating	d
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	5
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	5

In [34]:

```
sns.countplot(x='month', hue='type', data=df)
```

Out[34]:

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



In []: