

In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
import pandas as pd
df=pd.read_csv(r"C:\Users\Hello\Desktop\Netflic.csv")
```

In [137]:

```
# shape of the data
print('Rows in the data:',df.shape[0])
print('columns in the data:',df.shape[1])
```

Rows in the data: 8807  
columns in the data: 12

In [138]:

```
df.head()
```

Out[138]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

In [139]:

```
df.tail()
```

Out[139]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie	Zubaan	Mozes Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...

In [10]:

```
df.select_dtypes(include='object').describe()
```

```
# Description of the categorical Features
```

Out[10]:

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

In [140]:

```
#columns in data
```

```
df.columns
```

Out[140]:

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

In [141]:

```
# DataTypes of the Columns
```

```
df.dtypes
```

Out[141]:

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

In [181]:

```
# Statical Analysis
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 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        8807 non-null  object
4   cast            8807 non-null  object
5   country         8807 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: 894.5+ KB
```

In [142]:

```
#ALL Categorical features in Data
```

```
df.select_dtypes(include = 'object').columns
```

Out[142]:

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

In [143]:

```
df.select_dtypes(include='object').head()
```

Out[143]:

	show_id	type	title	director	cast	country	date_added	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

In [144]:

```
# Missing Value Detection
df.isnull().sum(),df.isnull().sum()*100/df.shape[0]
# here are IN the Director column there are 29% values are null...
```

Out[144]:

```
(show_id      0
type          0
title         0
director    2634
cast        825
country     831
date_added   10
release_year  0
rating        4
duration      3
listed_in     0
description   0
dtype: int64,
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 [145]:

```
# Stastical Summary
df.describe()

# Describe function describe the statical summary of all the numerical variable then it gives
# Count,mean,std,min,25%,50%,75%,max
```

Out[145]:

	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 [146]:

```
print('count of the type:',df['type'].value_counts()),df['type'].unique() # count of the type of the Moies
```

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

Out[146]:

```
(None, array(['Movie', 'TV Show'], dtype=object))
```

In [147]:

```
print('count of the title:',df['title'].value_counts()),df['title'].unique() # count and unique value for title
```

```
count of the title: Dick Johnson Is Dead      1
Ip Man 2      1
Hannibal Buress: Comedy Camisado      1
Turbo FAST      1
Masha's Tales      1
..
Love for Sale 2      1
ROAD TO ROMA      1
Good Time      1
Captain Underpants Epic Choice-o-Rama      1
Zubaan      1
Name: title, Length: 8807, dtype: int64
```

Out[147]:

```
(None,
 array(['Dick Johnson Is Dead', 'Blood & Water', 'Ganglands', ...,
       'Zombieland', 'Zoom', 'Zubaan'], dtype=object))
```

In [148]:

```
print('count of the director:',df['director'].value_counts()),df['director'].unique() # cout and unique val for the direcor
```

```
count of the director: Rajiv Chilaka      19
Raúl Campos, Jan Suter      18
Marcus Raboy      16
Suhas Kadav      16
Jay Karas      14
..
Raymie Muzquiz, Stu Livingston      1
Joe Menendez      1
Eric Bross      1
Will Eisenberg      1
Mozes Singh      1
Name: director, Length: 4528, dtype: int64
```

Out[148]:

```
(None,
 array(['Kirsten Johnson', nan, 'Julien Leclercq', ..., 'Majid Al Ansari',
       'Peter Hewitt', 'Mozes Singh'], dtype=object))
```

In [149]:

```
print('count of the case:',df['cast'].value_counts()),df['cast'].unique() # count adn unique val for cast
```

```
count of the case: David Attenborough
19
Vatsal Dubey, Julie Tejjwani, Rupa Bhimani, Jigna Bhardwaj, Rajesh Kava, Mousam, Swapnil
14
Samuel West
10
Jeff Dunham
7
David Spade, London Hughes, Fortune Feimster
6

..
Michael Peña, Diego Luna, Tenoch Huerta, Joaquin Cosio, José María Yazpik, Matt Letscher, Alyssa Diaz
1
Nick Lachey, Vanessa Lachey
1
Takeru Sato, Kasumi Arimura, Haru, Kentaro Sakaguchi, Takayuki Yamada, Kendo Kobayashi, Ken Yasuda, Arata Furuta, Suzuki Ma
tsuo, Koichi Yamadera, Arata Iura, Chikako Kaku, Kotaro Yoshida
1
Toyin Abraham, Sambasa Nzeribe, Chioma Chukwuka Akpotha, Chioma Omeruah, Chiwetelu Agu, Dele Odule, Femi Adebayo, Bayray Mc
Nwizu, Biodun Stephen
1
Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary, Meghna Malik, Malkeet Rauni, Anita Shabdish, Chittaranja
n Tripathy
1
Name: cast, Length: 7692, dtype: int64
```

Out[149]:

```
(None,
 array([nan,
        'Ama Qamata, Khosi Ngema, Gail Mabalone, Thabang Molaba, Dillon Windvogel, Natasha Thahane, Arno Greeff, Xolile Tsh
abalala, Getmore Sithole, Cindy Mahlangu, Ryle De Morny, Greteli Fincham, Sello Maake Ka-Ncube, Odwa Gwanya, Mekaila Mathy
s, Sandi Schultz, Duane Williams, Shamilla Miller, Patrick Mofokeng',
        'Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabiha Akkari, Sofia Lesaffre, Salim Kechiouche, Nouredine Farihi, Geer
t Van Rampelberg, Bakary Diombera',
        ...,
        'Jesse Eisenberg, Woody Harrelson, Emma Stone, Abigail Breslin, Amber Heard, Bill Murray, Derek Graf',
        'Tim Allen, Courteney Cox, Chevy Chase, Kate Mara, Ryan Newman, Michael Cassidy, Spencer Breslin, Rip Torn, Kevin Z
egers',
        'Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary, Meghna Malik, Malkeet Rauni, Anita Shabdish, Ch
ittaranjan Tripathy'],
      dtype=object))
```

In [150]:

```
print('count of the country:',df['country'].value_counts()),df['country'].unique(),len(df['country'].unique())
```

```
# there are 749 unique country in which movie was producing
```

```
United States, Czech Republic, Mauritius,
'Canada, South Africa', 'Austria', 'Mexico, Brazil',
'Germany, France', 'Mexico, United States',
'United Kingdom, France, Spain, United States',
'United States, Australia',
'United States, United Kingdom, France', 'United States, Russia',
'United States, United Kingdom, New Zealand',
'Australia, United Kingdom', 'Canada, Nigeria, United States',
'France, United States, United Kingdom, Canada',
'France, United Kingdom', 'India, United Kingdom',
'Canada, United States, Mexico',
'United Kingdom, Germany, United States',
'Czech Republic, United Kingdom, United States',
'China, United Kingdom', 'Italy, United Kingdom', 'China, Taiwan',
'United States, Brazil, Japan, Spain, India',
'United States, China, United Kingdom', 'Cameroon',
'Lebanon, Palestine, Denmark, Qatar', 'Japan, United States',
'Uruguay, Germany', 'Egypt, Saudi Arabia',
'United Kingdom, France, Poland, Germany, United States',
'Ireland, Switzerland, United Kingdom, France, United States',
...
```

In [151]:

```
print('cou of the rating:',df['rating'].value_counts()),df['rating'].unique()
```

```
cou of the rating: TV-MA      3207
TV-14      2160
TV-PG      863
R          799
PG-13      490
TV-Y7      334
TV-Y       307
PG          287
TV-G       220
NR          80
G          41
TV-Y7-FV    6
NC-17       3
UR          3
74 min      1
84 min      1
66 min      1
Name: rating, dtype: int64
```

Out[151]:

```
(None,
 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 [152]:

```
print('count of the description:',df['description'].value_counts()),df['description'].unique()
```

```
count of the description: Paranormal activity at a lush, abandoned property alarms a group eager to redevelop the site, but
the eerie events may not be as unearthly as they think.      4
Challenged to compose 100 songs before he can marry the girl he loves, a tortured but passionate singer-songwriter embarks
on a poignant musical journey.      3
A surly septuagenarian gets another chance at her 20s after having her photo snapped at a studio that magically takes 50 ye
ars off her life.      3
Multiple women report their husbands as missing but when it appears they are looking for the same man, a police officer tra
ces their cryptic connection.      3
Secrets bubble to the surface after a sensual encounter and an unforeseen crime entangle two friends and a woman caught bet
ween them.      2

..
Sent away to evade an arranged marriage, a 14-year-old begins a harrowing journey of sex work and poverty in the slums of A
ccra.      1
When his partner in crime goes missing, a small-time crook's life is transformed as he dedicates himself to raising the dau
ghter his friend left behind.      1
During 1962's Cuban missile crisis, a troubled math genius finds himself drafted to play in a U.S.-Soviet chess match - and
a deadly game of espionage.      1
A teen's discovery of a vintage Polaroid camera develops into a darker tale when she finds that whoever takes their photo w
ith it dies soon afterward.      1
A scrappy but poor boy worms his way into a tycoon's dysfunctional family, while facing his fear of music and the truth abo
ut his past.      1
Name: description, Length: 8775, dtype: int64
```

Out[152]:

```
(None,
 array(['As her father nears the end of his life, filmmaker Kirsten Johnson stages his death in inventive and comical ways
to help them both face the inevitable.',
       'After crossing paths at a party, a Cape Town teen sets out to prove whether a private-school swimming star is her
sister who was abducted at birth.',
       'To protect his family from a powerful drug lord, skilled thief Mehdi and his expert team of robbers are pulled int
o a violent and deadly turf war.',
       ...,
       'Looking to survive in a world taken over by zombies, a dorky college student teams with an urban roughneck and a p
air of grifter sisters.',
       'Dragged from civilian life, a former superhero must train a new crop of youthful saviors when the military preps f
or an attack by a familiar villain.',
       "A scrappy but poor boy worms his way into a tycoon's dysfunctional family, while facing his fear of music and the
truth about his past."],
      dtype=object))
```

In [153]:

```
df.head()
```

Out[153]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajilla, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

## Visual Analysis - Univariate, Bivariate after pre-processing of the data

In [154]:

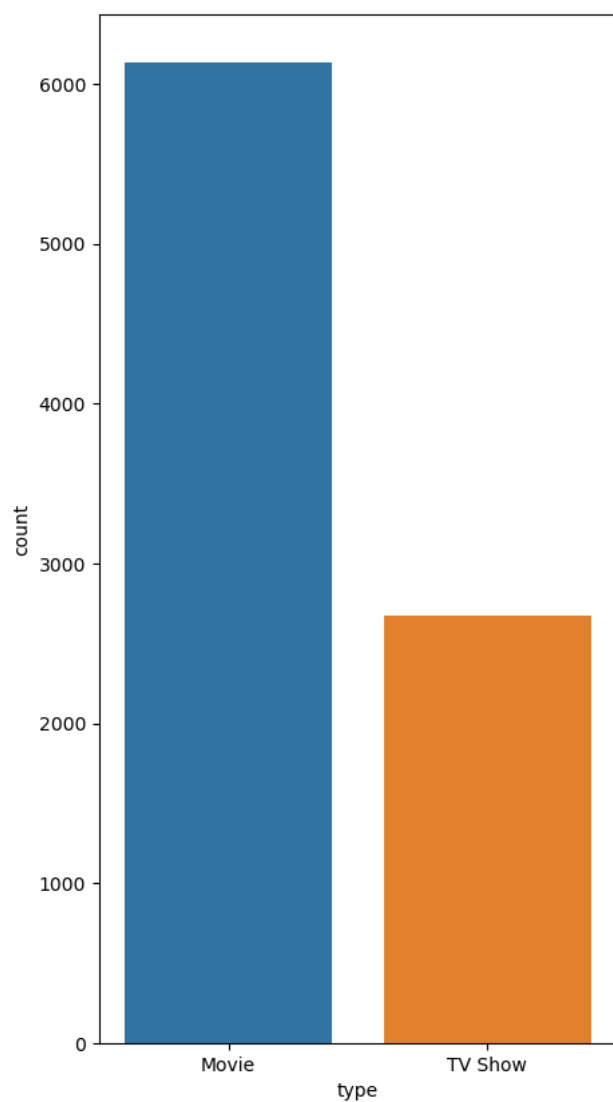
```
df.columns
```

Out[154]:

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

In [155]:

```
# Univariate Analysis
sns.countplot(df['type'])
plt.rcParams['figure.figsize'] = (5,5)
```



In [30]:

```
most_director_produced_movie = (df['director'].str.split(', ').explode().value_counts())

most_director_produced_movie
# there are multiple Names of the director which we most movie producer director
```

Out[30]:

```
Rajiv Chilaka      22
Jan Suter          21
Raúl Campos        19
Suhas Kadav        16
Marcus Raboy       16
..
Raymie Muzquiz     1
Stu Livingston     1
Joe Menendez       1
Eric Bross         1
Mozes Singh        1
Name: director, Length: 4993, dtype: int64
```



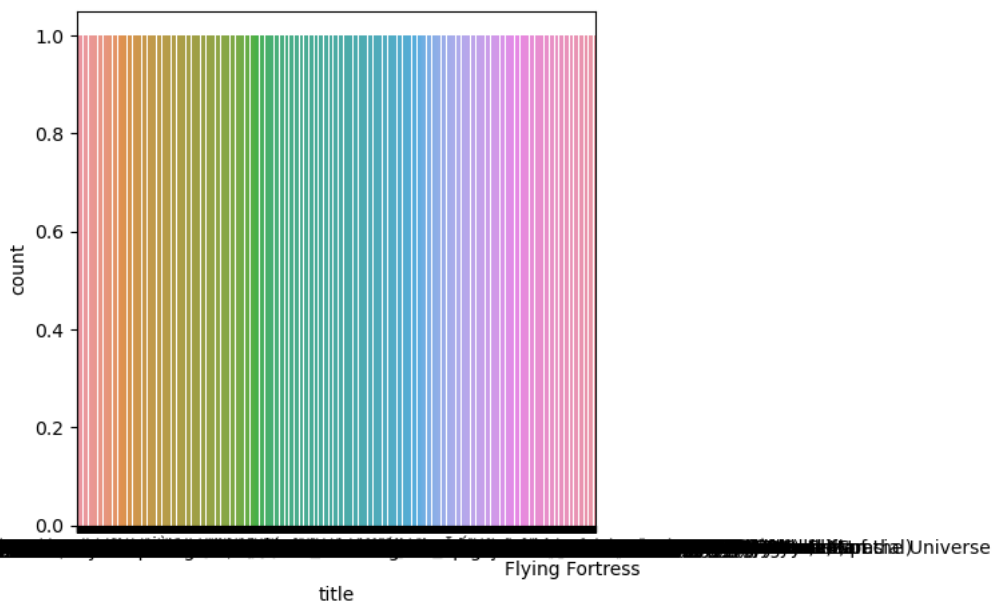
In [157]:

```
sns.countplot(df['title'], df['title'].value_counts())
```

```
# we cannot visualize the title of the movie by the graph
```

Out[157]:

```
(<AxesSubplot:xlabel='title', ylabel='count'>,
 Dick Johnson Is Dead      1
 Ip Man 2                  1
 Hannibal Buress: Comedy  1
 Turbo FAST                1
 Masha's Tales             1
 ..
 Love for Sale 2           1
 ROAD TO ROMA              1
 Good Time                 1
 Captain Underpants Epic  1
 Zubaan                    1
 Name: title, Length: 8807, dtype: int64)
```



In [20]:

```
top_producing_countries = (
    df['country']
    .str.split(',')
    .explode()
    .value_counts()
    .head(10)
)
top_producing_countries
```

Out[20]:

```
United States    3689
India            1046
United Kingdom   804
Canada           445
France           393
Japan            318
Spain            232
South Korea      231
Germany          226
Mexico           169
Name: country, dtype: int64
```

```
sns.countplot(df['country']) # there are multiple country which are not visualize by countplot
plt.pie(x = top_producing_countries.values, labels=top_producing_countries.index, autopct='% .2f%', data=df)
plt.show()
```

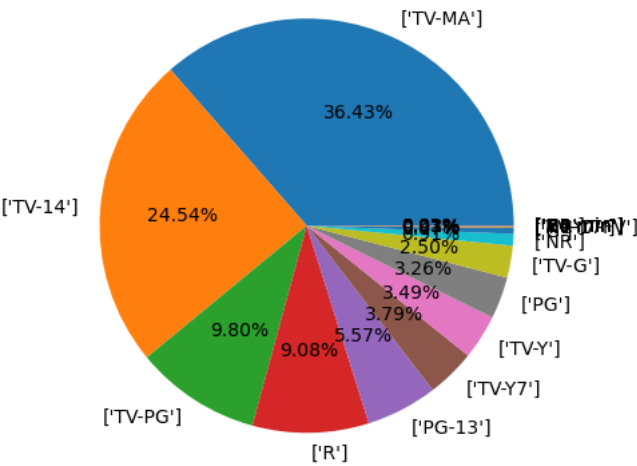


```
sns.countplot(df['rating'])
plt.rcParams['figure.figsize'] = (10,20)
```

# in the Countplot we can saw that the VTMA to rating is max another rating

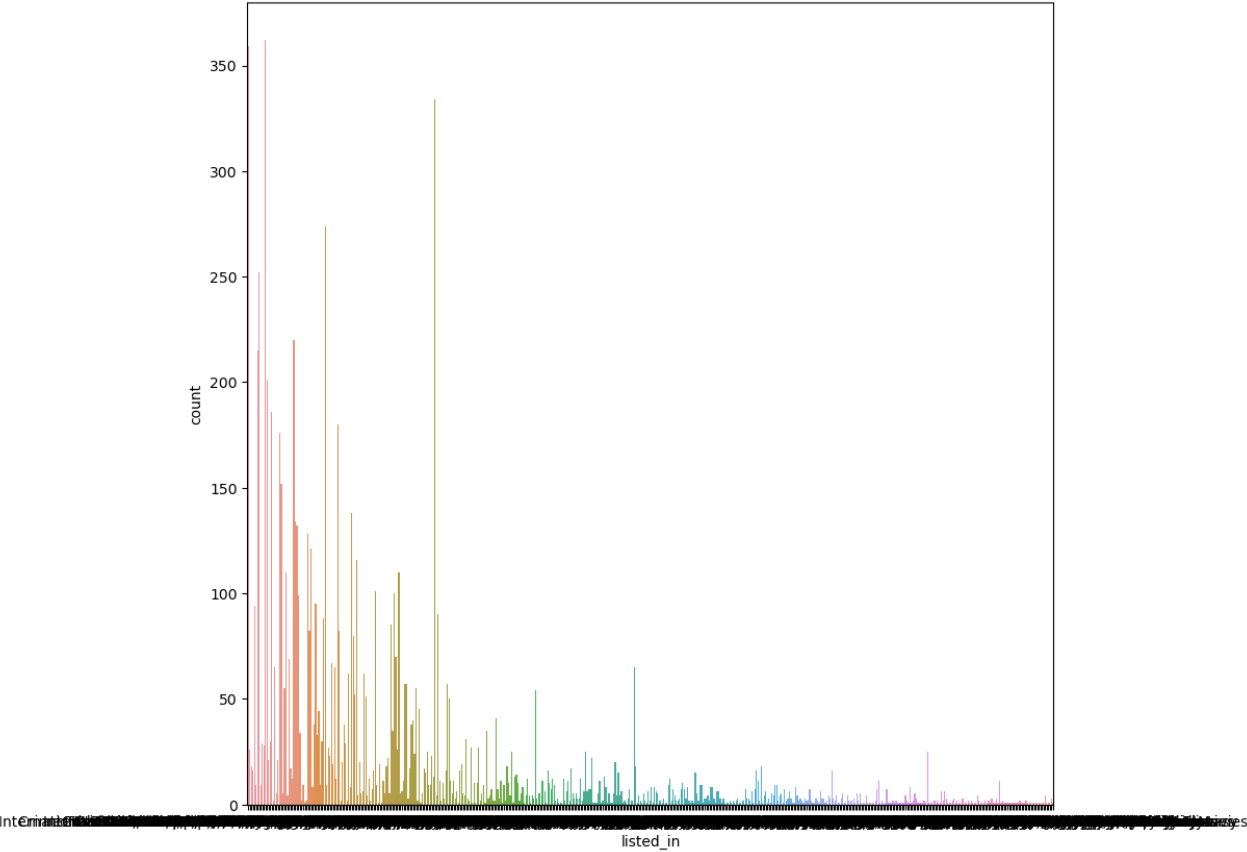


```
In [31]:
most_rating_movie = (df['rating'].str.split(', ').value_counts())
plt.pie(x = most_rating_movie.values,labels= most_rating_movie.index,autopct='%.2f%%')
plt.show()
```



```
In [160]:
plt.rcParams['figure.figsize'] = (10,10)
sns.countplot(df['listed_in']) # not visulaize by normal graph
```

Out[160]:  
<AxesSubplot:xlabel='listed\_in', ylabel='count'>

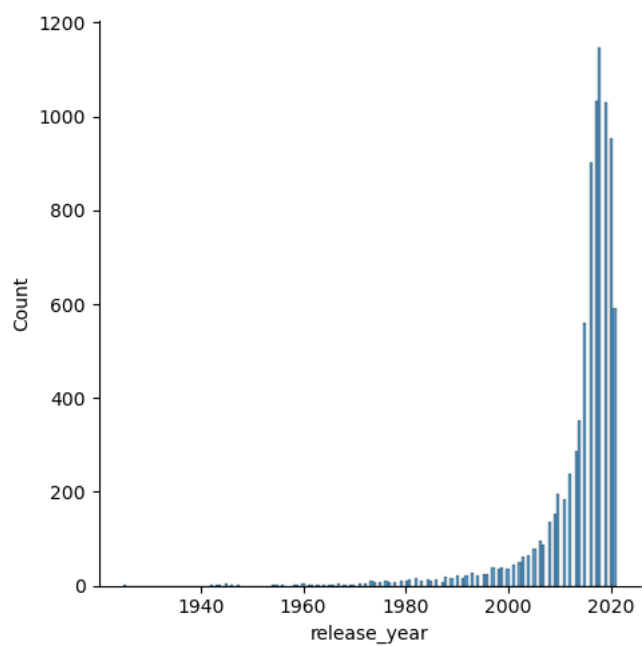


In [42]:

```
sns.displot(df['release_year'],df['release_year'].value_counts().sort_values(ascending=False))  
# after seeing the distplot we analaye that tmost of the movies released in 2018
```

Out[42]:

```
(<seaborn.axisgrid.FacetGrid at 0x1ebb30871f0>,  
2018    1147  
2017    1032  
2019    1030  
2020     953  
2016     902  
...  
1959         1  
1925         1  
1961         1  
1947         1  
1966         1  
Name: release_year, Length: 74, dtype: int64)
```

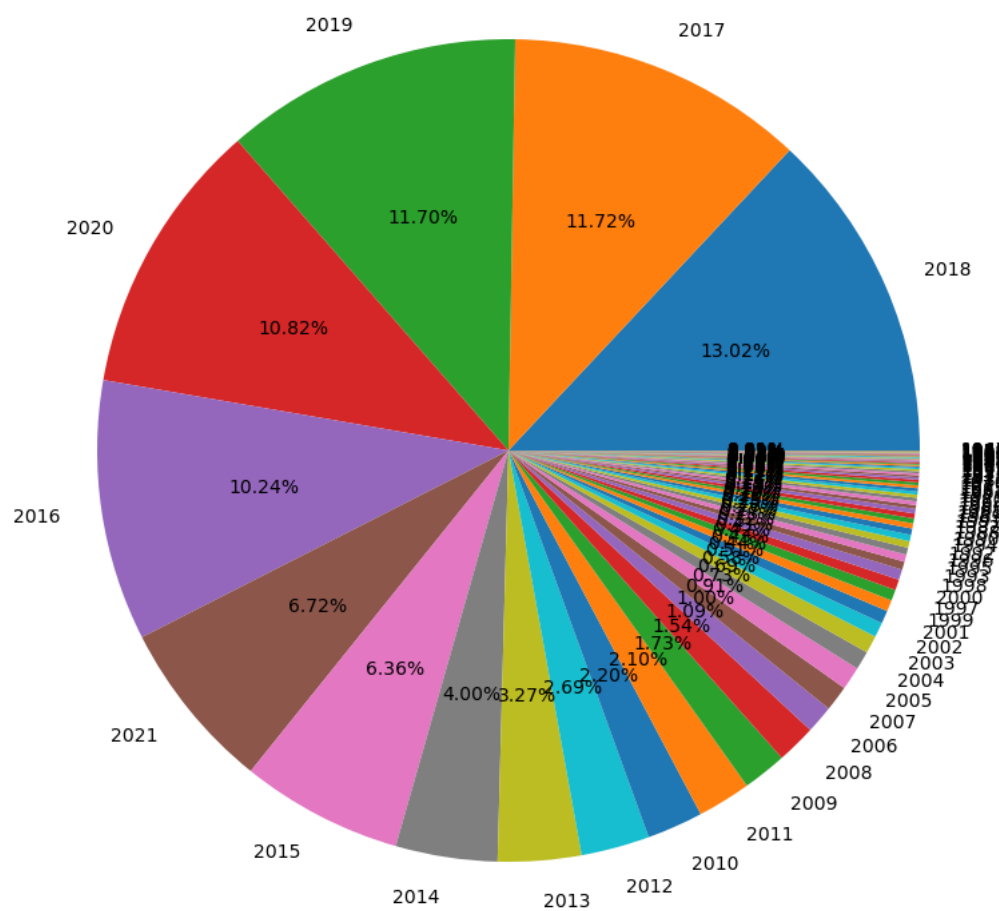


In [35]:

```

most_released_year = (df['release_year'].value_counts())
plt.pie(x = most_released_year, labels=most_released_year.index, autopct='%0.2f%%')
plt.show()
plt.rcParams['figure.figsize'] = (10,20)

```



In [162]:

```
sns.countplot(df['duration'], df['duration'].value_counts().sort_values(ascending=False))
```

*# we analyze that the most duration of the movie is 1 Season*

Out[162]:

(<AxesSubplot:xlabel='duration', ylabel='count'>,

1 Season 1793

2 Seasons 425

3 Seasons 199

90 min 152

94 min 146

...

167 min 1

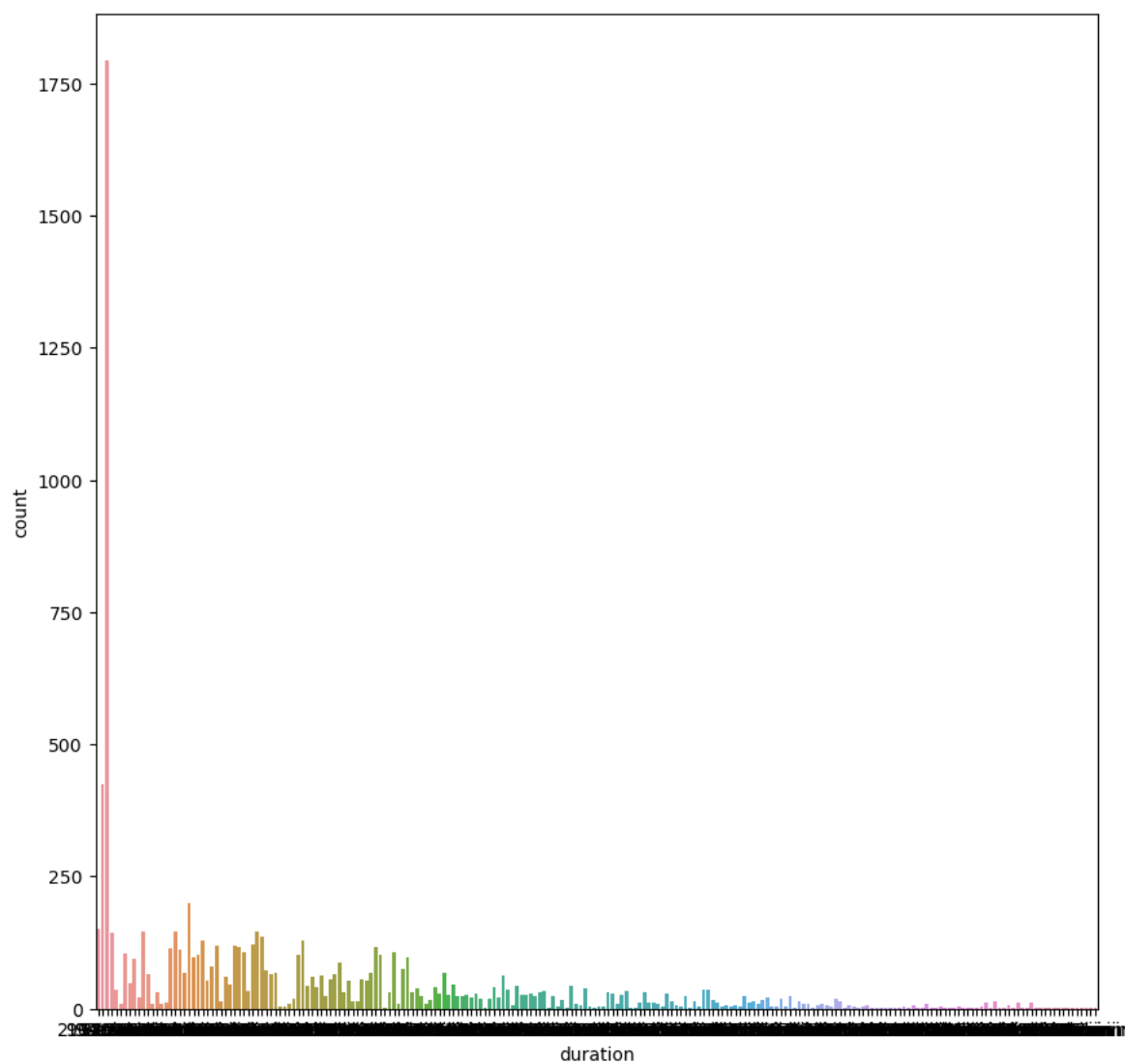
43 min 1

196 min 1

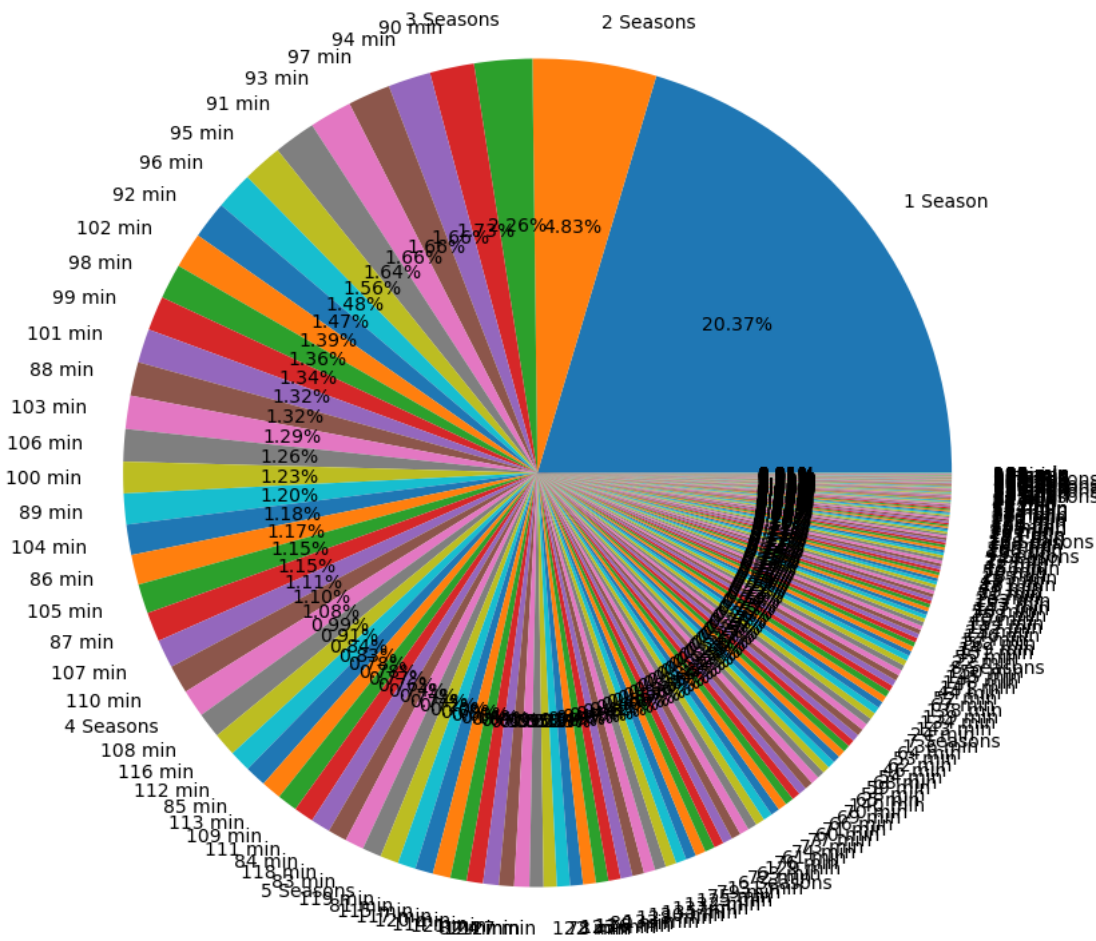
200 min 1

191 min 1

Name: duration, Length: 220, dtype: int64)



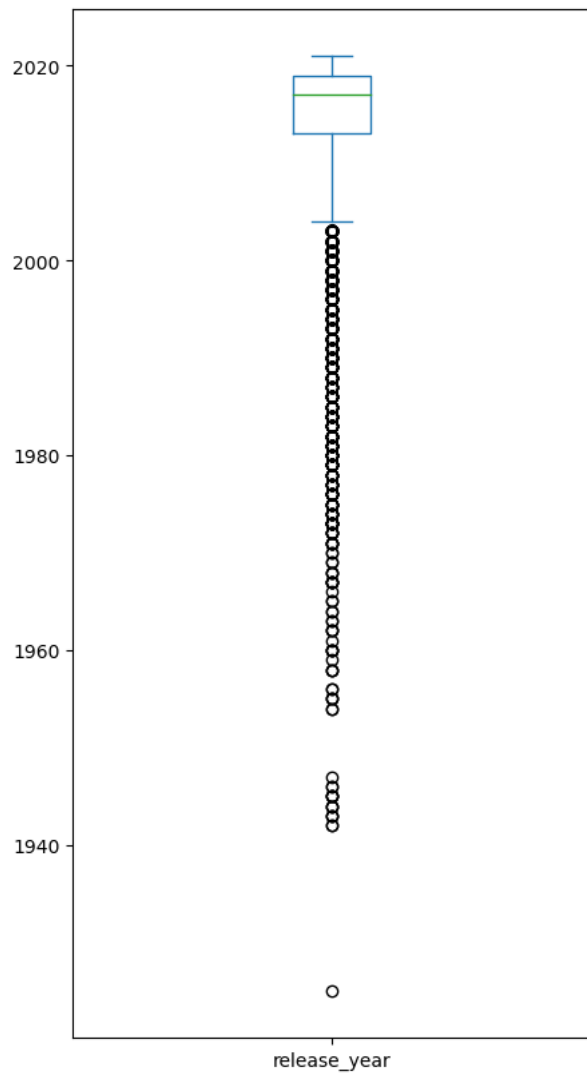
```
most_duration = (df['duration'].value_counts())
plt.pie(x = most_duration.values, labels=most_duration.index, autopct='%0.2f%%')
plt.show()
```



In [7]:

```
df.select_dtypes(include=np.number).plot(kind='box')  
plt.rcParams['figure.figsize'] = (5,10)
```

*# In the our data we have only one numerical columns that was release\_year*





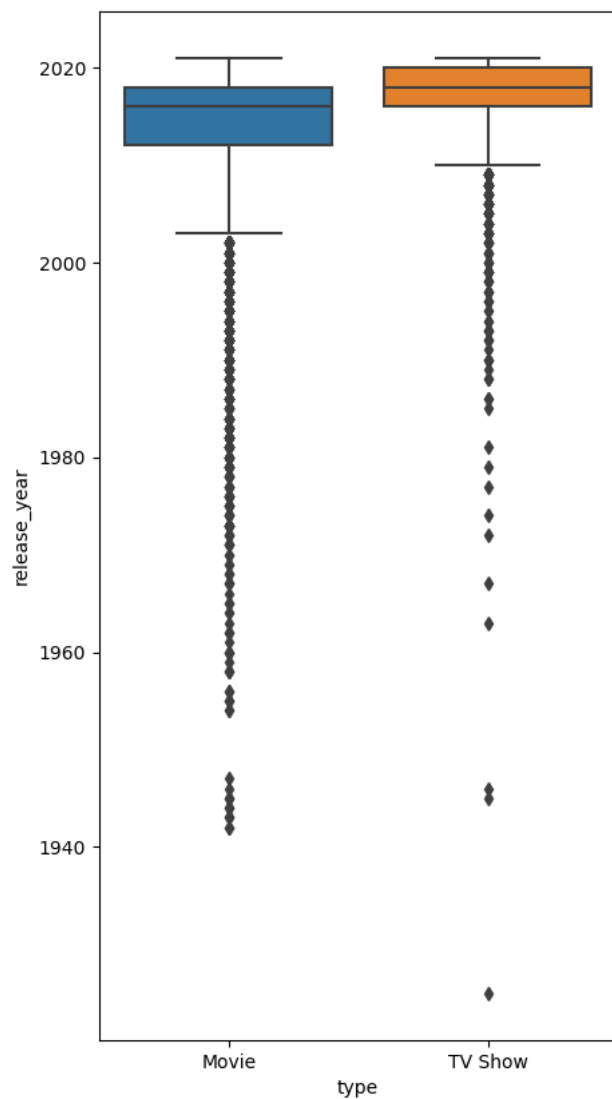
In [6]:

```
sns.boxplot(df.type,df.release_year,data=df)
```

```
# we analyze that year wise TVshow and movies Far values
```

Out[6]:

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

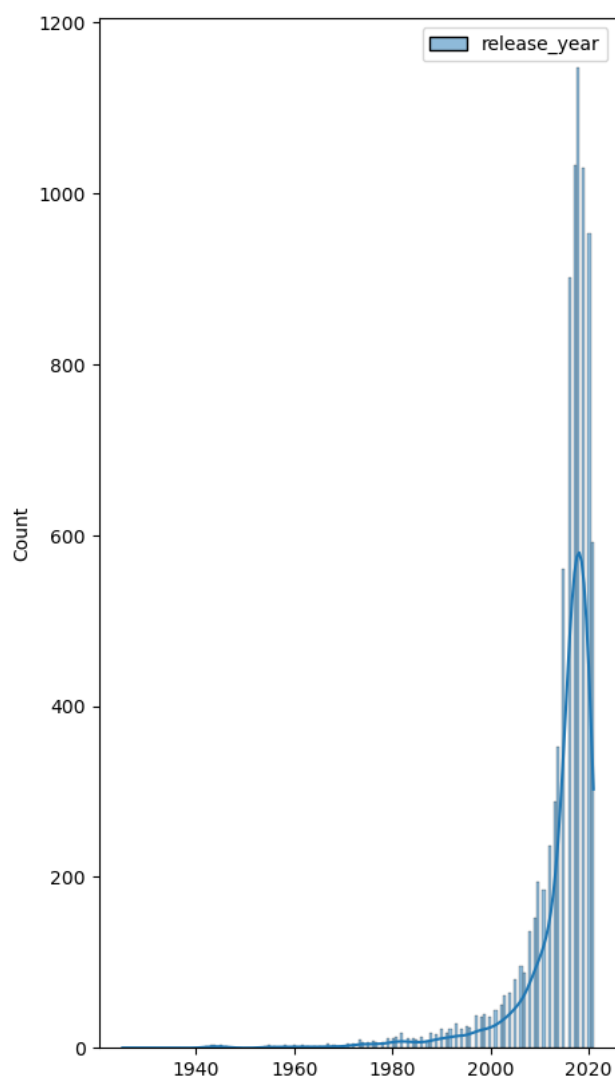


In [165]:

```
sns.histplot(df,kde=True)
```

Out[165]:

<AxesSubplot:ylabel='Count'>

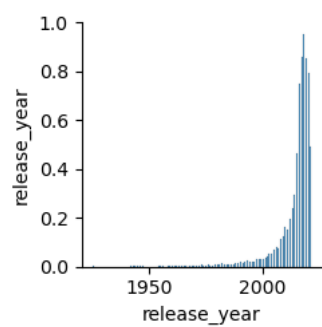


In [166]:

```
sns.pairplot(df)
```

Out[166]:

<seaborn.axisgrid.PairGrid at 0x1e70e7685b0>



## NULL VALUES PERCENTAGE

In [176]:

```
df['director'] = df['director'].fillna(df['director'].mode()[0])
df['country'] = df['country'].fillna(df['country'].mode()[0])
df['cast'] = df['cast'].fillna(df['cast'].mode()[0])
```

In [178]:

```
df.drop_duplicates(inplace=True)
```

In [179]:

```
round(df.isnull().sum()*100/df.shape[0],2)
```

Out[179]:

```
show_id      0.00
type         0.00
title        0.00
director     0.00
cast         0.00
country      0.00
date_added   0.11
release_year 0.00
rating       0.05
duration     0.03
listed_in    0.00
description  0.00
dtype: float64
```

## OUTLIER TREATMENT

In [5]:

```
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)

IQR = Q3 - Q1

upper_limit = Q3 + 1.5*IQR
Lower_Limit = Q1 - 1.5*IQR
```

In [169]:

```
print('Upper_limit:',upper_limit,'Lowe_Limit:',Lower_Limit)
# IN OUR DATASET WE HAVE ONLY ONE NUMERICAL FEATURES AND THERE ARE UPPER_LIMIT AND LOWER_LIMIT ARE VALUES WHICH ARE VERY FAR
#FROM THE MEAN VALUES
```

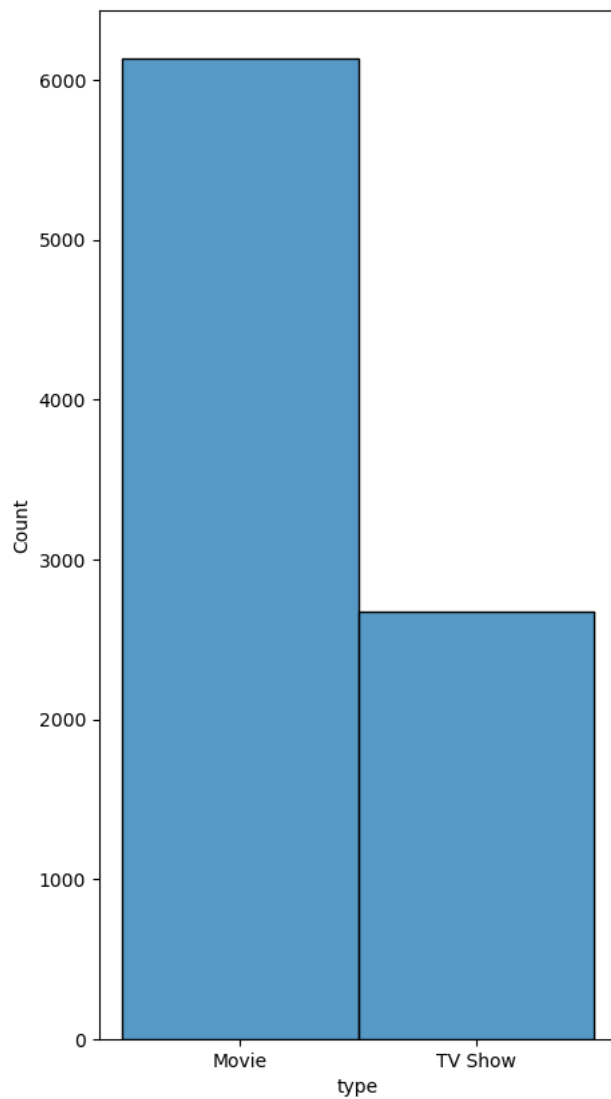
```
Upper_limit: release_year    2028.0
dtype: float64 Lowe_Limit: release_year    2004.0
dtype: float64
```

In [170]:

```
sns.histplot(df['type'])
```

Out[170]:

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



In [172]:

```
df
```

Out[172]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajjila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...

8807 rows × 12 columns

FEATURE ENGINEERING

In [188]:

```
sns.palplot(['#221f1f', '#b20710', '#e50914','#f5f5f1'])

plt.title("Netflix brand palette ",loc='left',fontfamily='serif',fontsize=15,y=1.2)
plt.show()
x=df.groupby(['type'])['type'].count()
y=len(df)
r=((x/y)).round(2)

mf_ratio = pd.DataFrame(r).T
```

Netflix brand palette



In [9]:

```
# Question : 6
df.columns
```

Out[9]:

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

In [37]:

```
months_name = (
    pd.to_datetime(df['date_added'])
    .dt.month_name()
)
months_name
```

Out[37]:

```
0      September
1      September
2      September
3      September
4      September
...
8802   November
8803      July
8804   November
8805   January
8806      March
Name: date_added, Length: 8807, dtype: object
```

In [40]:

```
content_type_and_upload_month = (
    pd.concat([df['type'], months_name], axis=1)
    .rename(columns={'date_added': 'month_added'})
    .value_counts()
)
content_type_and_upload_month
```

Out[40]:

```
type      month_added
Movie      July         565
           April         550
           December      547
           January        546
           October        545
           March          529
           September      519
           August         519
           November       498
           June           492
           May            439
           February       382
TV Show   December       266
           July           262
           September      251
           August         236
           June           236
           October        215
           April          214
           March          213
           November       207
           May            193
           January        192
           February       181
dtype: int64
```

In [ ]: