```
In [1]:  import numpy as np
         import pandas as pd
         data=pd.read_csv('netflix.net_public_assets_assets_000_000_940_original_net
         import seaborn as sns
         import matplotlib.pyplot as plt
```

```
In [2]:  data.head()
```

Out[2]:

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

Understand the columns

```
In [3]:  data.info()
```

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

In [4]: `data.shape`

Out[4]: `(8807, 12)`

In [5]: `data.describe()`

Out[5]:

	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 [6]: `data.describe(include='object')`

Out[6]:

	show_id	type	title	director	cast	country	date_added	rating	durat
count	8807	8807	8807	6173	7982	7976	8797	8803	88
unique	8807	2	8807	4528	7692	748	1767	17	
top	s1	Movie	Dick Johnson Is Dead	Rajiv Chilaka	David Attenborough	United States	January 1, 2020	TV-MA	1 Sea
freq	1	6131	1	19	19	2818	109	3207	1

Handling Missing Values

In [7]: `data.isnull()`

Out[7]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	dur
0	False	False	False	False	True	False	False	False	False	
1	False	False	False	True	False	False	False	False	False	
2	False	False	False	False	False	True	False	False	False	
3	False	False	False	True	True	True	False	False	False	
4	False	False	False	True	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	
8802	False	False	False	False	False	False	False	False	False	
8803	False	False	False	True	True	True	False	False	False	
8804	False	False	False	False	False	False	False	False	False	
8805	False	False	False	False	False	False	False	False	False	
8806	False	False	False	False	False	False	False	False	False	

8807 rows × 12 columns

In [8]:
```python
data.isnull().sum()/len(data)*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 [8]:
```python
data['director'].isnull().value_counts()
```

Out[8]:
```
False    6173
True     2634
Name: director, dtype: int64
```

In [9]:
```python
filtered_data=data[data['director'].isnull()][['type','director']]
filtered_data['type']
```

Out[9]:
```
1       TV Show
3       TV Show
4       TV Show
10      TV Show
14      TV Show
         ...
8795    TV Show
8796    TV Show
8797    TV Show
8800    TV Show
8803    TV Show
Name: type, Length: 2634, dtype: object
```
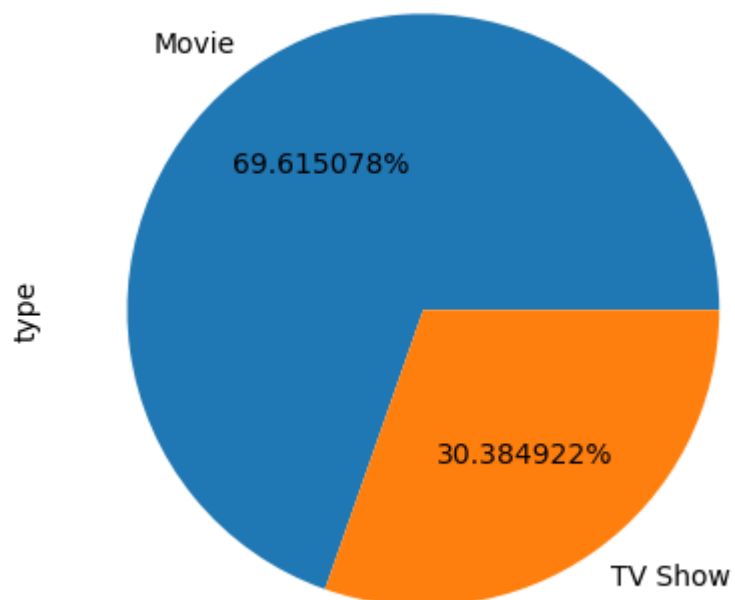
In [10]:
```python
data['type'].value_counts(normalize=True)*100
```

Out[10]:
```
Movie      69.615079
TV Show    30.384921
Name: type, dtype: float64
```

In [11]: `data['type'].value_counts().plot(kind='pie',autopct='%1f%%')`

Out[11]: `<Axes: ylabel='type'>`

In [12]: `data.head()`

Out[12]:

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

In [13]:
```
constraint=data['cast'].apply(lambda x:str(x).split(',')).tolist
constraint
```

Out[13]:
```
<bound method IndexOpsMixin.tolist of 0
[nan]
1       [Ama Qamata,  Khosi Ngema,  Gail Mabalane,  Th...
2       [Sami Bouajila,  Tracy Gotoas,  Samuel Jouy,  ...
3                                                  [nan]
4       [Mayur More,  Jitendra Kumar,  Ranjan Raj,  Al...
                              ...
8802    [Mark Ruffalo,  Jake Gyllenhaal,  Robert Downe...
8803                                               [nan]
8804    [Jesse Eisenberg,  Woody Harrelson,  Emma Ston...
8805    [Tim Allen,  Courteney Cox,  Chevy Chase,  Kat...
8806    [Vicky Kaushal,  Sarah-Jane Dias,  Raaghav Cha...
Name: cast, Length: 8807, dtype: object>
```

In [14]:
```
New_data=data['duration'].apply(lambda x:str(x).split(',')).tolist()
```

In [15]:
```
data.loc[1:5,'show_id':'country']
```

Out[15]:

	show_id	type	title	director	cast	country
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN

In [16]: `data`

Out[16]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-	India	March 2, 2019	2015

	show_id	type	title	director	cast	country	date_added	release_year
					Jane Dias, Raaghav Chanan...			

8807 rows × 12 columns

In [17]:
```python
data.sort_values('release_year',ascending=False).head(5)
```

Out[17]:

	show_id	type	title	director	cast	country	date_added	release_year
693	s694	Movie	Ali & Ratu Ratu Queens	Lucky Kuswandi	Iqbaal Ramadhan, Nirina Zubir, Asri Welas, Tik...	NaN	June 17, 2021	2021
781	s782	Movie	Black Holes	The Edge of All We Know	Peter Galison	NaN	NaN	June 2, 2021	2021
762	s763	Movie	Sweet & Sour	Lee Kae-byeok	Jang Ki-yong, Chae Soo-bin, Jung Soo-jung	South Korea	June 4, 2021	2021
763	s764	TV Show	Sweet Tooth	NaN	Nonso Anozie, Christian Convery, Adeel Akhtar,...	United States	June 4, 2021	2021
764	s765	Movie	Trippin' with the Kandasamys	Jayan Moodley	Jailoshini Naidoo, Maeshni Naicker, Madhushan ...	South Africa	June 4, 2021	2021

In [18]:
```python
def encode(X):
    if X=='Movie':
        return 1
    else:
        return 0
```

In [19]:
```python
data['Type_Mapping']=data['type'].apply(encode)
data
```

Out[19]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-	India	March 2, 2019	2015

show_id	type	title	director	cast	country	date_added	release_year
				Jane Dias, Raaghav Chanan...			

8807 rows × 13 columns

```
In [20]: data['director'].nunique()
```

Out[20]: 4528

```
In [21]: data.groupby('director').ngroups
```

Out[21]: 4528

```
In [22]: data.groupby('director').get_group('David Fincher')
```

Out[22]:

	show_id	type	title	director	cast	country	date_added	release_year	ra
600	s601	Movie	The Game	David Fincher	Michael Douglas, Sean Penn, Deborah Kara Unger...	United States	July 1, 2021	1997	
1595	s1596	Movie	MANK	David Fincher	Gary Oldman, Amanda Seyfried, Charles Dance, L...	United States	December 4, 2020	2020	
7701	s7702	Movie	Panic Room	David Fincher	Jodie Foster, Forest Whitaker, Dwight Yoakam, ...	United States	August 1, 2019	2002	
8320	s8321	Movie	The Girl with the Dragon Tattoo	David Fincher	Daniel Craig, Rooney Mara, Christopher Plummer...	United States, Sweden, Norway	January 5, 2021	2011	
8511	s8512	Movie	The Social Network	David Fincher	Jesse Eisenberg, Andrew Garfield, Justin Timbe...	United States	April 1, 2020	2010	P
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	

```
In [23]: data.drop('Type_Mapping',axis=1)
```

Out[23]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-	India	March 2, 2019	2015

	show_id	type	title	director	cast	country	date_added	release_year
					Jane Dias, Raaghav Chanan...			

8807 rows × 12 columns

In [24]: `data.groupby('director').get_group('David Fincher')['title'].count()`

Out[24]: 6

Most Productive Director

In [25]: `data['director'].value_counts()`

Out[25]:
```
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
Mozez Singh                        1
Name: director, Length: 4528, dtype: int64
```

In [26]: `data['director'].fillna('Unknown',inplace=True)`

In [27]: `data['director']`

Out[27]:
```
0         Kirsten Johnson
1                 Unknown
2         Julien Leclercq
3                 Unknown
4                 Unknown
              ...
8802        David Fincher
8803              Unknown
8804      Ruben Fleischer
8805         Peter Hewitt
8806          Mozez Singh
Name: director, Length: 8807, dtype: object
```
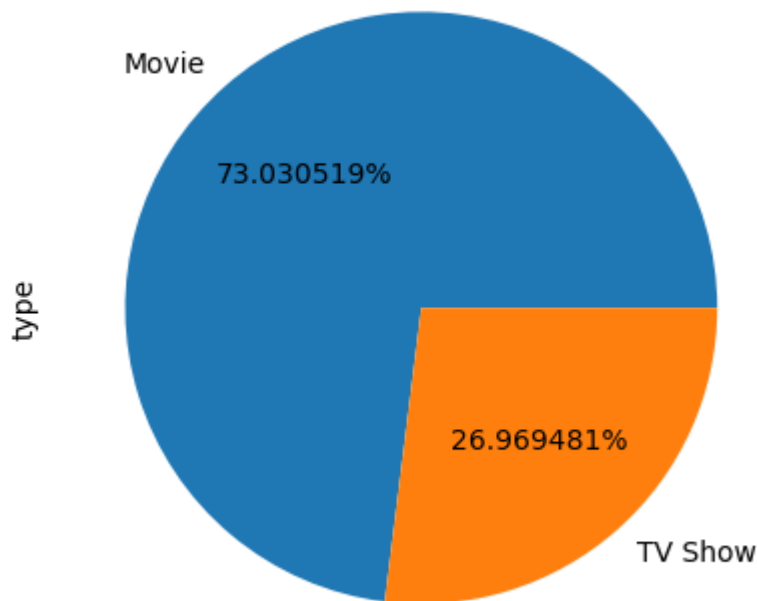
In [28]: `data_type=data['type'].unique`
`data_type`

Out[28]:
```
<bound method Series.unique of 0          Movie
1         TV Show
2         TV Show
3         TV Show
4         TV Show
           ...
8802        Movie
8803      TV Show
8804        Movie
8805        Movie
8806        Movie
Name: type, Length: 8807, dtype: object>
```

In [29]:
```python
data[data['country']=='United States']['type'].value_counts().plot(kind='pie
```

Out[29]:
```
<Axes: ylabel='type'>
```


According to the data above 73 % people watch Movie and almost 27% people watch TV shows in United States.

In [30]:
```python
data[data['type']=='TV Show']['country'].value_counts().sort_values(ascendir
```

Out[30]:
```
United States      760
United Kingdom     213
Japan              169
South Korea        158
India               79
Taiwan              68
Canada              59
France              49
Spain               48
Australia           48
Name: country, dtype: int64
```

Above data consist top 10 countries Publishing TV shows on netflix.

In [31]:
```python
data[data['type']=='Movie']['country'].value_counts().sort_values(ascending=
```

```
Out[31]:  United States    2058
          India            893
          United Kingdom   206
          Canada           122
          Spain            97
          Egypt            92
          Nigeria          86
          Indonesia        77
          Turkey           76
          Japan            76
          Name: country, dtype: int64
```

Above data consist top 10 countries Publishing Movie on netflix.

```
In [32]:  data['yrs_max']=data['release_year'].max()
          data['yrs_min']=data['release_year'].min()
          data
```

Out[32]:

	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	Unknown	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	Unknown	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	Unknown	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	Unknown	NaN	NaN	July 1, 2019	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-	India	March 2, 2019	2015

show_id	type	title	director	cast	country	date_added	release_year
				Jane Dias, Raaghav Chanan...			

8807 rows × 15 columns

In [33]:   `data.drop('yrs_max',axis=1)`

Out[33]:

	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	Unknown	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	Unknown	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	Unknown	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	Unknown	NaN	NaN	July 1, 2019	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-	India	March 2, 2019	2015

show_id	type	title	director	cast	country	date_added	release_year
				Jane Dias, Raaghav Chanan...			

8807 rows × 14 columns

In [34]: `data.drop(['yrs_min','yrs_max'],axis=1)`

Out[34]:

	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	Unknown	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	Unknown	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	Unknown	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	Unknown	NaN	NaN	July 1, 2019	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-	India	March 2, 2019	2015

show_id	type	title	director	cast	country	date_added	release_year
				Jane Dias, Raaghav Chanan...			

8807 rows × 13 columns

Visual Analysis - Univariate, Bivariate after Pre-processing of the Data

In [35]:
```python
sns.distplot(data['release_year'].dropna(), kde=False, bins=30)
plt.title('Distribution of Release Years')
plt.show()
```
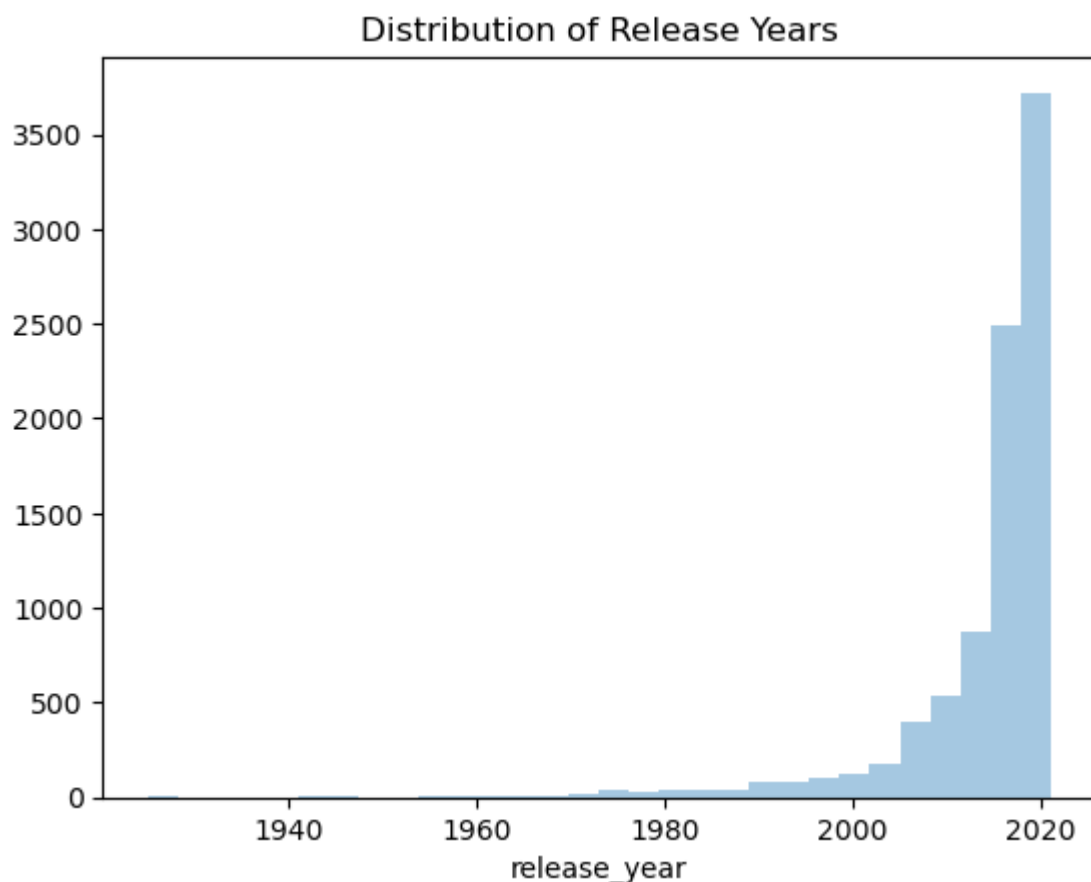
/var/folders/fr/xk4r1t0n0kqdxrwjlb8fd4n80000gn/T/ipykernel_15028/411058115
2.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
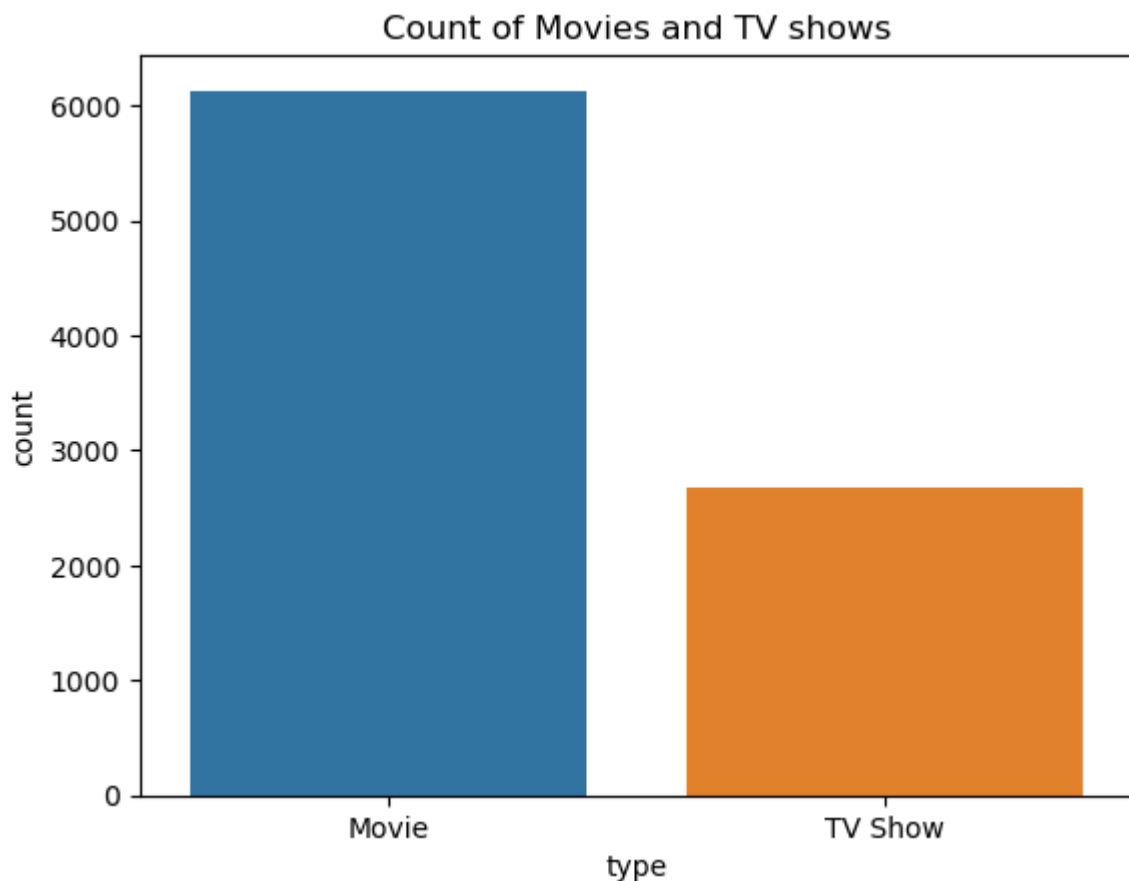similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(data['release_year'].dropna(), kde=False, bins=30)

Distribution of Release Years

from past 2000 number of movie release has been increase over the year

In [36]:
```python
sns.countplot(x='type',data=data)
plt.title('Count of Movies and TV shows')
plt.show()
```


Count of Movies and TV shows

In [37]:
```python
#unnesting cast columns
df_cast=data['cast'].apply(lambda x:str(x).split(',')).tolist()
cast=pd.DataFrame(df_cast, index=data['title'])
Cast=cast.stack()
Cast=pd.DataFrame(Cast.reset_index())
Cast.rename(columns={0:'Cast'},inplace=True)
Cast.drop(['level_1'],axis=1,inplace=True)
Cast
```

Out[37]:

	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 [38]:
```python
# unnesting listed_in columns
Genre=data['listed_in'].apply(lambda x:str(x).split(',')).tolist()
Genre=pd.DataFrame(Genre,index=data['title'])
Genre=Genre.stack()
Genre=pd.DataFrame(Genre.reset_index())
Genre.rename(columns={0:'Genre'},inplace=True)
Genre.drop(['level_1'],axis=1,inplace=True)
Genre
```

Out[38]:

	title	Genre
0	Dick Johnson Is Dead	Documentaries
1	Blood & Water	International TV Shows
2	Blood & Water	TV Dramas
3	Blood & Water	TV Mysteries
4	Ganglands	Crime TV Shows
...	...	...
19318	Zoom	Children & Family Movies
19319	Zoom	Comedies
19320	Zubaan	Dramas
19321	Zubaan	International Movies
19322	Zubaan	Music & Musicals

19323 rows × 2 columns

In [39]:
```python
Fam_Gen=Genre.groupby(['Genre']).agg({'title':'nunique'}).reset_index().sor
```

In [40]:
```python
Fam_Gen
```

Out[40]:

	Genre	title
13	International Movies	2624
49	Dramas	1600
44	Comedies	1210
37	Action & Adventure	859
47	Documentaries	829
...	...	...
59	Romantic Movies	3
62	Spanish-Language TV Shows	2
70	TV Sci-Fi & Fantasy	1
55	LGBTQ Movies	1
63	Sports Movies	1

73 rows × 2 columns

In [41]:
```python
data['cast'].replace([np.nan],['Unknown'],inplace=True)
```

In [42]:
```python
data['director'].replace([np.nan],['Unknown'],inplace=True)
```

In [43]:
```python
data.isnull().sum()/len(data) * 100
```

Out[43]:
```
show_id         0.000000
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        0.034064
listed_in       0.000000
description     0.000000
Type_Mapping    0.000000
yrs_max         0.000000
yrs_min         0.000000
dtype: float64
```

In [44]:
```python
data['rating'].unique().tolist()
```

```
Out[44]:   ['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']
```

```
In [45]:   mode=data['rating'].mode()
           mode
```

```
Out[45]:   0    TV-MA
           Name: rating, dtype: object
```

```
In [46]:   data['rating'].replace([np.nan],['Unknown'],inplace=True)
```

```
In [47]:   data.isnull().sum()/ len(data) * 100
```

```
Out[47]:   show_id          0.000000
           type             0.000000
           title            0.000000
           director         0.000000
           cast             0.000000
           country          9.435676
           date_added       0.113546
           release_year     0.000000
           rating           0.000000
           duration         0.034064
           listed_in        0.000000
           description      0.000000
           Type_Mapping     0.000000
           yrs_max          0.000000
           yrs_min          0.000000
           dtype: float64
```

```
In [48]:   data.groupby(['rating']).agg({"title":"nunique"}).reset_index().sort_values(
```

Out[48]:

	rating	title
11	TV-MA	3207
9	TV-14	2160
12	TV-PG	863
8	R	799
7	PG-13	490
14	TV-Y7	334
13	TV-Y	307
6	PG	287
10	TV-G	220
5	NR	80
3	G	41
15	TV-Y7-FV	6
17	Unknown	4
4	NC-17	3
16	UR	3
1	74 min	1
2	84 min	1
0	66 min	1

In [49]:
```python
data_movie=data[data['type']=='Movie']
data_movie
```

Out[49]:

	show_id	type	title	director	cast	country	date_added	release_
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Unknown	United States	September 25, 2021	
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	NaN	September 24, 2021	
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	
12	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	
...	...	...	...	...	...	...	...	
8801	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	

6131 rows × 15 columns

```
In [50]:  data_shows=data[data['type']=='TV Show']
          data_shows
```

Out[50]:

	show_id	type	title	director	cast	country	date_added	release_yea
1	s2	TV Show	Blood & Water	Unknown	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	202
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	202
3	s4	TV Show	Jailbirds New Orleans	Unknown	Unknown	NaN	September 24, 2021	202
4	s5	TV Show	Kota Factory	Unknown	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	202
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	202
...	...	...	...	...	...	...	...	..
8795	s8796	TV Show	Yu-Gi-Oh! Arc-V	Unknown	Mike Liscio, Emily Bauer, Billy Bob Thompson, ...	Japan, Canada	May 1, 2018	201
8796	s8797	TV Show	Yunus Emre	Unknown	Gökhan Atalay, Payidar Tüfekçioglu, Baran Akbu...	Turkey	January 17, 2017	2016
8797	s8798	TV Show	Zak Storm	Unknown	Michael Johnston, Jessica Gee-George, Christin...	United States, France, South Korea, Indonesia	September 13, 2018	2016
8800	s8801	TV Show	Zindagi Gulzar Hai	Unknown	Sanam Saeed, Fawad Khan, Ayesha Omer, Mehreen ...	Pakistan	December 15, 2016	2012
8803	s8804	TV Show	Zombie Dumb	Unknown	Unknown	NaN	July 1, 2019	2018

show_id	type	title	director	cast	country	date_added	release_yea

2676 rows × 15 columns

In [51]:
```python
data_movie['duration'].apply(lambda x: str(x).split(',')).tolist
```

Out[51]:
```
<bound method IndexOpsMixin.tolist of 0          [90 min]
6          [91 min]
7          [125 min]
9          [104 min]
12         [127 min]
              ...
8801        [96 min]
8802       [158 min]
8804        [88 min]
8805        [88 min]
8806       [111 min]
Name: duration, Length: 6131, dtype: object>
```

In [52]:
```python
data_India_movie=data_movie[data_movie['country']=='India']
data_India_movie
```

Out[52]:

	show_id	type	title	director	cast	country	date_added	release_year
24	s25	Movie	Jeans	S. Shankar	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...	India	September 21, 2021	1998
105	s106	Movie	Angamaly Diaries	Lijo Jose Pellissery	Antony Varghese, Reshma Rajan, Binny Rinky Ben...	India	September 5, 2021	2017
114	s115	Movie	Anjaam	Rahul Rawail	Madhuri Dixit, Shah Rukh Khan, Tinnu Anand, Jo...	India	September 2, 2021	1994
116	s117	Movie	Dhanak	Nagesh Kukunoor	Krrish Chhabria, Hetal Gada, Vipin Sharma, Gul...	India	September 2, 2021	2015
118	s119	Movie	Gurgaon	Shanker Raman	Akshay Oberoi, Pankaj Tripathi, Ragini Khanna,...	India	September 2, 2021	2017
...	...	...	...	...	...	...	...	...
8772	s8773	Movie	Yamla Pagla Deewana 2	Sangeeth Sivan	Dharmendra, Sunny Deol, Bobby Deol, Neha Sharm...	India	May 1, 2017	2013
8773	s8774	Movie	Yanda Kartavya Aahe	Kedar Shinde	Ankush Choudhary, Smita Shewale, Mohan Joshi, ...	India	January 1, 2018	2006
8798	s8799	Movie	Zed Plus	Chandra Prakash Dwivedi	Adil Hussain, Mona Singh, K.K. Raina, Sanjay M...	India	December 31, 2019	2014
8799	s8800	Movie	Zenda	Avadhoot Gupte	Santosh Juvekar, Siddharth Chandekar, Sachit P...	India	February 15, 2018	2009
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015

893 rows × 15 columns

In [53]:
```python
data_India_dur=data_India_movie['duration'].str.split(expand=True)
data_India_dur
```

Out[53]:

	0	1
24	166	min
105	128	min
114	143	min
116	114	min
118	106	min
...	...	...
8772	147	min
8773	151	min
8798	131	min
8799	120	min
8806	111	min

893 rows × 2 columns

In [54]:
```python
data_India_dur.rename(columns={0:'Duration_min'},inplace=True)
data_India_dur.drop([1],axis=1,inplace=True)
data_India_dur
```

Out[54]:

	Duration_min
24	166
105	128
114	143
116	114
118	106
...	...
8772	147
8773	151
8798	131
8799	120
8806	111

893 rows × 1 columns

In [55]:
```python
data_India_dur.astype(int).mean().round(2)
```

Out[55]:
```
Duration_min    126.92
dtype: float64
```