

```
In [1]: 1 # The dataset provided belongs to a popular entertainment streaming service in the world
2 # Since there is no data related to revenue we begin with the assumption that it has a profitable revenue
3 # The problem statement that we aim to answer below will be insightful and to see what can be done better
4 # In terms of content, what is popular among the audience? Movies or Series?
5 # What is the global reach of this platform?
6 # In which country is the most content released?
```

```
In [2]: 1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
```

```
In [3]: 1 #Load the dataset using pandas Library for EDA
2 nf = pd.read_csv('C:/Scaler-MODULES/Netflix/Netflix.csv')
```

```
In [4]: 1 #check the initial data to see the preview of how the data looks
2 #the imported data gives the 2d list in the form of table which is called 'dataframe'
3 #Each and every column is called 'series'
4 nf.head(5)
```

Out[4]:

	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, film...
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 I...

```
In [5]: 1 #we identify and explore the different variables and datatypes available in the dataset
2 #This gives the information about overall count of non null data,datatypes of the each column
3 nf.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 [6]: 1 #Returns the dimension of the data in the form of tuple
2 #we see that there are 8807 rows and 12 columns from the dataset provided
3 nf.shape
```

Out[6]: (8807, 12)

```
In [7]: 1 # keys return the columns in the form of explicit indexes and each key has values associated with it
2 # which are the rows for each column
3 nf.keys()
```

Out[7]: Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
 'release_year', 'rating', 'duration', 'listed_in', 'description'],
 dtype='object')

```
In [30]: 1 # Now we can analyse the basic metrics of the given data
2 # This gives the statistical summary of the data for int and floating values
3 nf.describe()
```

```
Out[30]: 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]:

```
1 # This will return summary of stastistical table for all the columns including object columns
2 nf.describe(include='all')
```

Out[7]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
count	8807	8807	8807	6173	7982	7976	8797	8807.000000	8803	8804	8807	8807
unique	8807	2	8807	4528	7692	748	1767	NaN	17	220	514	8775
top	s1	Movie	Dick Johnson Is Dead	Rajiv Chilaka	David Attenborough	United States	January 1, 2020	NaN	TV-MA	1 Season	Dramas, International Movies	Paranormal activity at a lush, abandoned prop...
freq	1	6131	1	19	19	2818	109	NaN	3207	1793	362	4
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2014.180198	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	8.819312	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1925.000000	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2013.000000	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2017.000000	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2019.000000	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2021.000000	NaN	NaN	NaN	NaN

In [8]:

```
1 #Includes only numerical columns for aggregation
2 nf.describe(include=[np.number])
```

Out[8]:

	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 [9]:

```
1 #This will return transposed summary of the data where columns are now the index and rows represent measures
2 # include='object' parameter ensure that only object columns are included
3 # we see that out of 8807 rows, movies account to 6131
4 nf.describe(include='object').T
```

Out[9]:

	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 prop...	4

```
In [10]: 1 # analysing null values
2 # isnull() or isna() both provide same functionality where it returns the missing or null values
3 # and can be used interchangeably with each other. here sum() will count all the true values which are the null values
4 nf.isna().sum()
```

```
Out[10]: 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
```

```
In [35]: 1 # this will return the number of missing values in percentage.
2 # we see 29.9% of values missing in director column and 9% in cast and country columns
3 (nf.isna().sum()/nf.shape[0]) *100
```

```
Out[35]: 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 [36]: 1 # The data contains movies which were released from 1942 to 2021
2 nf.release_year.nunique(),nf.release_year.unique()
```

```
Out[36]: (74,
array([2020, 2021, 1993, 2018, 1996, 1998, 1997, 2010, 2013, 2017, 1975,
       1978, 1983, 1987, 2012, 2001, 2014, 2002, 2003, 2004, 2011, 2008,
       2009, 2007, 2005, 2006, 1994, 2015, 2019, 2016, 1982, 1989, 1990,
       1991, 1999, 1986, 1992, 1984, 1980, 1961, 2000, 1995, 1985, 1976,
       1959, 1988, 1981, 1972, 1964, 1945, 1954, 1979, 1958, 1956, 1963,
       1970, 1973, 1925, 1974, 1960, 1966, 1971, 1962, 1969, 1977, 1967,
       1968, 1965, 1946, 1942, 1955, 1944, 1947, 1943], dtype=int64))
```

```
In [11]: 1 # We have to preprocess the data before we can analyse using visual plots
2 # From the data we can see that director,cast and listed_in columns has multiple values
3 #This will transform the cast column where in each value is unique
4
5 constraint = nf['cast'].apply(lambda x: str(x).split(', ')).tolist()
6 nf_new = pd.DataFrame(constraint, index=nf['title'])
7 nf_new= nf_new.stack()
8 nf_new = pd.DataFrame(nf_new)
9 nf_new.reset_index(inplace=True)
10 nf_new = nf_new[['title',0]]
11 nf_new.columns =['title','cast']
12 nf_new
```

Out[11]:

	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 [12]: 1 #This will transform the director column where in each value is unique
2
3 constraint1 = nf['director'].apply(lambda x: str(x).split(', ')).tolist()
4 nf_d = pd.DataFrame(constraint1, index=nf['title'])
5 nf_d= nf_d.stack()
6 nf_d = pd.DataFrame(nf_d)
7 nf_d.reset_index(inplace=True)
8 nf_d = nf_d[['title',0]]
9 nf_d.columns =['title','director']
10 nf_d
```

Out[12]:

	title	director
0	Dick Johnson Is Dead	Kirsten Johnson
1	Blood & Water	nan
2	Ganglands	Julien Leclercq
3	Jailbirds New Orleans	nan
4	Kota Factory	nan
...
9607	Zodiac	David Fincher
9608	Zombie Dumb	nan
9609	Zombieland	Ruben Fleischer
9610	Zoom	Peter Hewitt
9611	Zubaan	Mozez Singh

9612 rows × 2 columns

```
In [13]: 1 #This will transform the country column where in each value is unique
2
3 constraint2 = nf['country'].apply(lambda x: str(x).split(', ')).tolist()
4 nf_c = pd.DataFrame(constraint2,index=nf['title'])
5 nf_c= nf_c.stack()
6 nf_c = pd.DataFrame(nf_c)
7 nf_c.reset_index(inplace=True)
8 nf_c = nf_c[['title',0]]
9 nf_c.columns =['title','country']
10 nf_c
```

Out[13]:

	title	country
0	Dick Johnson Is Dead	United States
1	Blood & Water	South Africa
2	Ganglands	nan
3	Jailbirds New Orleans	nan
4	Kota Factory	India
...
10840	Zodiac	United States
10841	Zombie Dumb	nan
10842	Zombieland	United States
10843	Zoom	United States
10844	Zubaan	India

10845 rows × 2 columns

```
In [14]: 1 #This will transform the listed_in column where in each value is unique
2
3 constraint3 = nf['listed_in'].apply(lambda x: str(x).split(', ')).tolist()
4 nf_l = pd.DataFrame(constraint3,index=nf['title'])
5 nf_l= nf_l.stack()
6 nf_l = pd.DataFrame(nf_l)
7 nf_l.reset_index(inplace=True)
8 nf_l = nf_l[['title',0]]
9 nf_l.columns =['title','listed_in']
10 nf_l
```

Out[14]:

	title	listed_in
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 [15]: 1 #we have to merge all the four dataframe to create new frmae which will be used to merge with o
2 nf_merge1 = nf_new.merge(nf_d)
3 nf_merge2 = nf_c.merge(nf_l)
4 nf_merge = nf_merge1.merge(nf_merge2)
5 nf_merge
```

Out[15]:

	title	cast	director	country	listed_in
0	Dick Johnson Is Dead	nan	Kirsten Johnson	United States	Documentaries
1	Blood & Water	Ama Qamata	nan	South Africa	International TV Shows
2	Blood & Water	Ama Qamata	nan	South Africa	TV Dramas
3	Blood & Water	Ama Qamata	nan	South Africa	TV Mysteries
4	Blood & Water	Khosi Ngema	nan	South Africa	International TV Shows
...
201986	Zubaan	Anita Shabdish	Mozez Singh	India	International Movies
201987	Zubaan	Anita Shabdish	Mozez Singh	India	Music & Musicals
201988	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	Dramas
201989	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	International Movies
201990	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	Music & Musicals

201991 rows × 5 columns

```
In [16]: 1 # merge with original dataset to have unique rows all through the dataset
2 nf_final = pd.merge(nf,nf_merge,on=['title'],how='inner')
3 nf_final
```

Out[16]:

	show_id	type	title	director_x	cast_x	country_x	date_added	release_year	rating	duration	listed_in_x	descripti
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Nan	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As father ne the enc his l filmr
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	At cross paths e party Cape To
2	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	At cross paths e party Cape To
3	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	At cross paths e party Cape To
4	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	At cross paths e party Cape To
...
201986	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrap but pi boy wor his way i a t
201987	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrap but pi boy wor his way i a t
201988	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrap but pi boy wor his way i a t
201989	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrap but pi boy wor his way i a t
201990	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrap but pi boy wor his way i a t

201991 rows x 16 columns

```
In [18]: 1 # we need to remove the old columns and rename the new columns with names from original dataset
2 nf_f= nf_final.drop(columns=['director_x','cast_x','country_x','listed_in_x'])
3 nf_f=nf_f.rename(columns={'cast_y':'cast','director_y':'director','country_y':'country','listed_in_y':'l
4 ordered= ['show_id','type','title', 'director', 'cast', 'country', 'date_added', 'release_year', 'rating'
5 nf_f=nf_f[ordered]
6 nf_f
```

Out[18]:

	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, film...
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
2	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...
3	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Mysteries	After crossing paths at a party, a Cape Town t...
4	s2	TV Show	Blood & Water	nan	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
...
201986	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201987	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...
201988	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Dramas	A scrappy but poor boy worms his way into a ty...
201989	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201990	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...

201991 rows × 12 columns

In [22]: 1 nf_f.shape

Out[22]: (201991, 12)

```
In [19]: 1 nf_f.loc[nf_f.duplicated()]
2 nf_f
```

Out[19]:

	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, filmmaker...
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
2	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...
3	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Mysteries	After crossing paths at a party, a Cape Town t...
4	s2	TV Show	Blood & Water	nan	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
...
201986	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201987	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...
201988	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Dramas	A scrappy but poor boy worms his way into a ty...
201989	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201990	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...

201991 rows × 12 columns



In [20]: 1 nf_f.loc[nf_f.duplicated(subset=['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added', 'rating', 'duration', 'listed_in', 'descript'])]

Out[20]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descript
39336	s1632	Movie	Rust Creek	Jen McGowan	Micah Hauptman	United States	November 30, 2020	2018	R	108 min	Thrillers	A wrong turn in two become fight fo
88474	s3719	Movie	Blood Will Tell	Miguel Cohan	Oscar Martinez	Argentina	June 21, 2019	2019	TV-MA	113 min	Dramas	Far patria E begin:unre aft
88475	s3719	Movie	Blood Will Tell	Miguel Cohan	Oscar Martinez	Argentina	June 21, 2019	2019	TV-MA	113 min	Independent Movies	Far patria E begin:unre aft

In [21]:

```

1 #we delete the duplicates based on the combination of all the columns combined
2 nf_du=nf_f.drop_duplicates(subset=['show_id','type','title', 'director', 'cast', 'country', 'date_added']
3 nf_du

```

Out[21]:

	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	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
2	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...
3	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Mysteries	After crossing paths at a party, a Cape Town t...
4	s2	TV Show	Blood & Water	nan	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
...
201986	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201987	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...
201988	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Dramas	A scrappy but poor boy worms his way into a ty...
201989	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201990	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...

201936 rows × 12 columns

```
In [22]: 1 #reset the index and setting the drop=True will remove the previous index column
2 nf_du.reset_index(drop=True,inplace =True)
3 nf_du
```

Out[22]:

	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, film...
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
2	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...
3	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Mysteries	After crossing paths at a party, a Cape Town t...
4	s2	TV Show	Blood & Water	nan	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
...
201931	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201932	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...
201933	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Dramas	A scrappy but poor boy worms his way into a ty...
201934	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201935	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...

201936 rows × 12 columns



```
In [23]: 1 #replacing the nan values with NA(Not Available)
2 nf_du['director']=nf_du['director'].replace('nan', 'NA')
```

C:\Users\Prashanth K M\AppData\Local\Temp\ipykernel_30124\4290520587.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
nf_du['director']=nf_du['director'].replace('nan', 'NA')

```
In [24]: 1 #replacing the nan values with NA(Not Available)
2 nf_du['cast']=nf_du['cast'].replace('nan', 'NA')
3 nf_du['country']=nf_du['country'].replace('nan', 'NA')
4 nf_du
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	dtype
1	1	TV Show	Blood & Water	NA	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Mysteries	A cross paths part Cape To	
2	2	TV Show	Blood & Water	NA	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	A cross paths part Cape To	
3	3	TV Show	Blood & Water	NA	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Mysteries	A cross paths part Cape To	
4	4	TV Show	Blood & Water	NA	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	A cross paths part Cape To	
...	
201931	201931	s8807	Movie	Zubaan	Mozes Sime	Anita Choudhary	March 2, 2018	2015	TV-14	111 min	International Movies	A scrub boy wo	

```
In [25]: 1 #changing date format of 'date_added' column
2 # nf_du['date_added']= pd.to_datetime(nf_du['date_added'])
3
4 (nf_du.isna().sum()/nf_du.shape[0])*100
5 #since these values constitute less than 1% we can choose to drop the rows.
```

```
Out[25]: show_id      0.000000
type        0.000000
title       0.000000
director    0.000000
cast        0.000000
country     0.000000
date_added  0.078243
release_year 0.000000
rating      0.033179
duration    0.001486
listed_in   0.000000
description 0.000000
dtype: float64
```

In [26]:

```

1 # nf_du.info()
2 nf_du = nf_du.dropna(subset=['date_added', 'rating', 'duration'])
3 nf_du.reset_index(drop=True, inplace=True)
4 nf_du

```

Out[26]:

	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	NA	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	NA	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
2	s2	TV Show	Blood & Water	NA	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...
3	s2	TV Show	Blood & Water	NA	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Mysteries	After crossing paths at a party, a Cape Town t...
4	s2	TV Show	Blood & Water	NA	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
...
201703	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201704	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...
201705	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Dramas	A scrappy but poor boy worms his way into a ty...
201706	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
201707	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...

201708 rows × 12 columns



```
In [27]: 1 # extracting month from the date added column
2 # changing date format of 'date_added' column
3 nf_du['date_added']= pd.to_datetime(nf_du['date_added'], format='mixed')
4 nf_du['month'] = nf_du['date_added'].dt.month_name(locale = 'English')
5 # nf_du(inplace=True)
6 nf_du.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 201708 entries, 0 to 201707
Data columns (total 13 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   show_id     201708 non-null   object  
 1   type        201708 non-null   object  
 2   title       201708 non-null   object  
 3   director    201708 non-null   object  
 4   cast        201708 non-null   object  
 5   country     201708 non-null   object  
 6   date_added  201708 non-null   datetime64[ns]
 7   release_year 201708 non-null   int64  
 8   rating      201708 non-null   object  
 9   duration    201708 non-null   object  
 10  listed_in   201708 non-null   object  
 11  description 201708 non-null   object  
 12  month       201708 non-null   object  
dtypes: datetime64[ns](1), int64(1), object(11)
memory usage: 20.0+ MB
```

```
C:\Users\Prashanth K M\AppData\Local\Temp\ipykernel_30124\1383451698.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

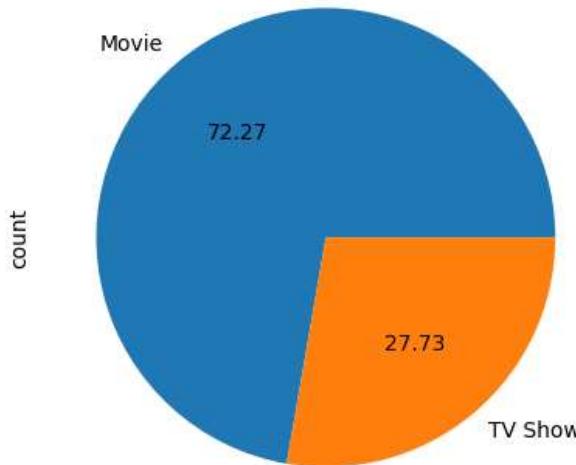
```
nf_du['date_added']= pd.to_datetime(nf_du['date_added'], format='mixed')
C:\Users\Prashanth K M\AppData\Local\Temp\ipykernel_30124\1383451698.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
nf_du['month'] = nf_du['date_added'].dt.month_name(locale = 'English')
```

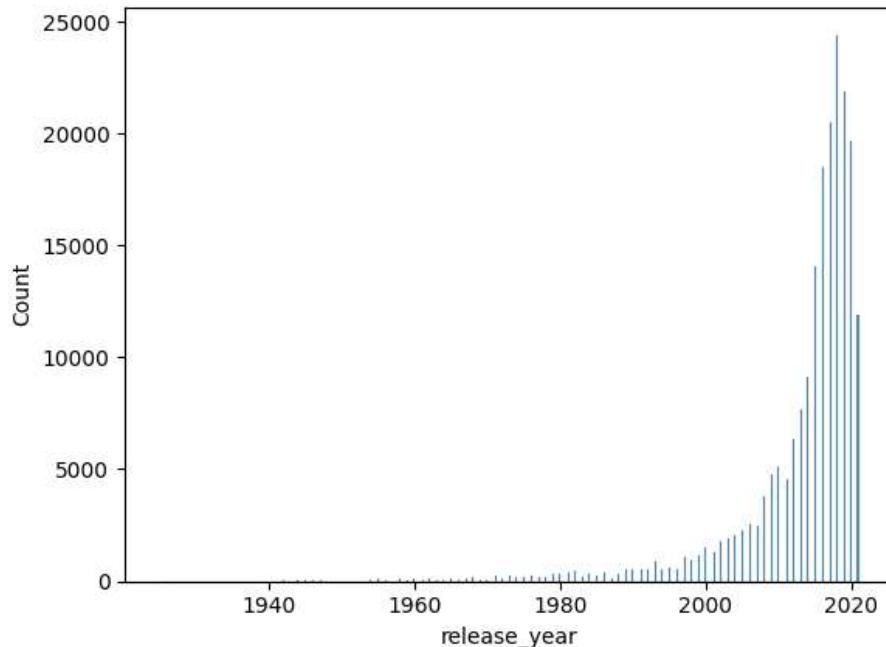
```
In [28]: 1 nf_du['type'].value_counts().plot(kind='pie', autopct='%.2f')
```

```
Out[28]: <Axes: ylabel='count'>
```



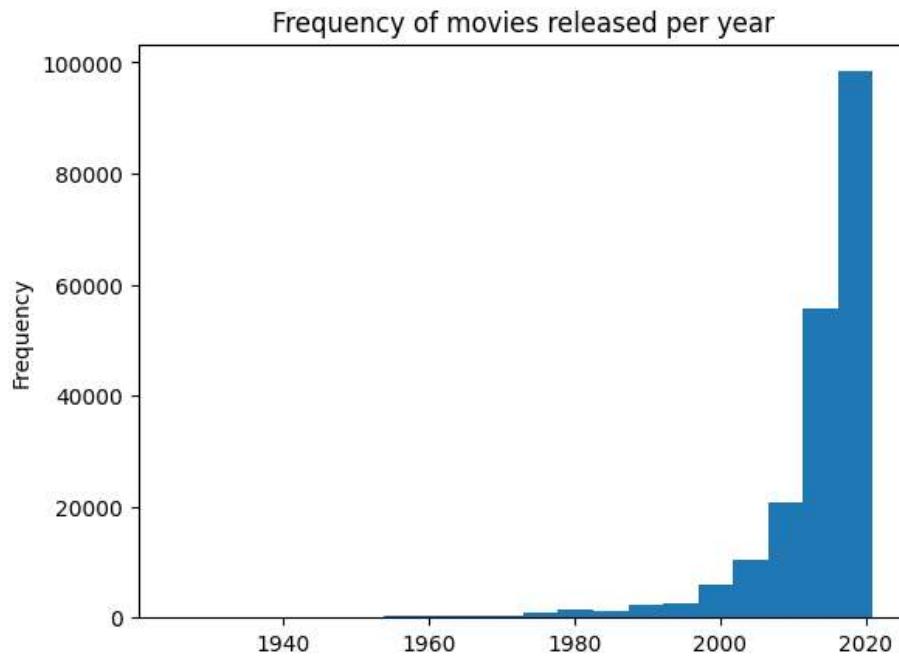
```
In [33]: 1 sns.histplot(data=nf_du,x='release_year')
```

```
Out[33]: <Axes: xlabel='release_year', ylabel='Count'>
```



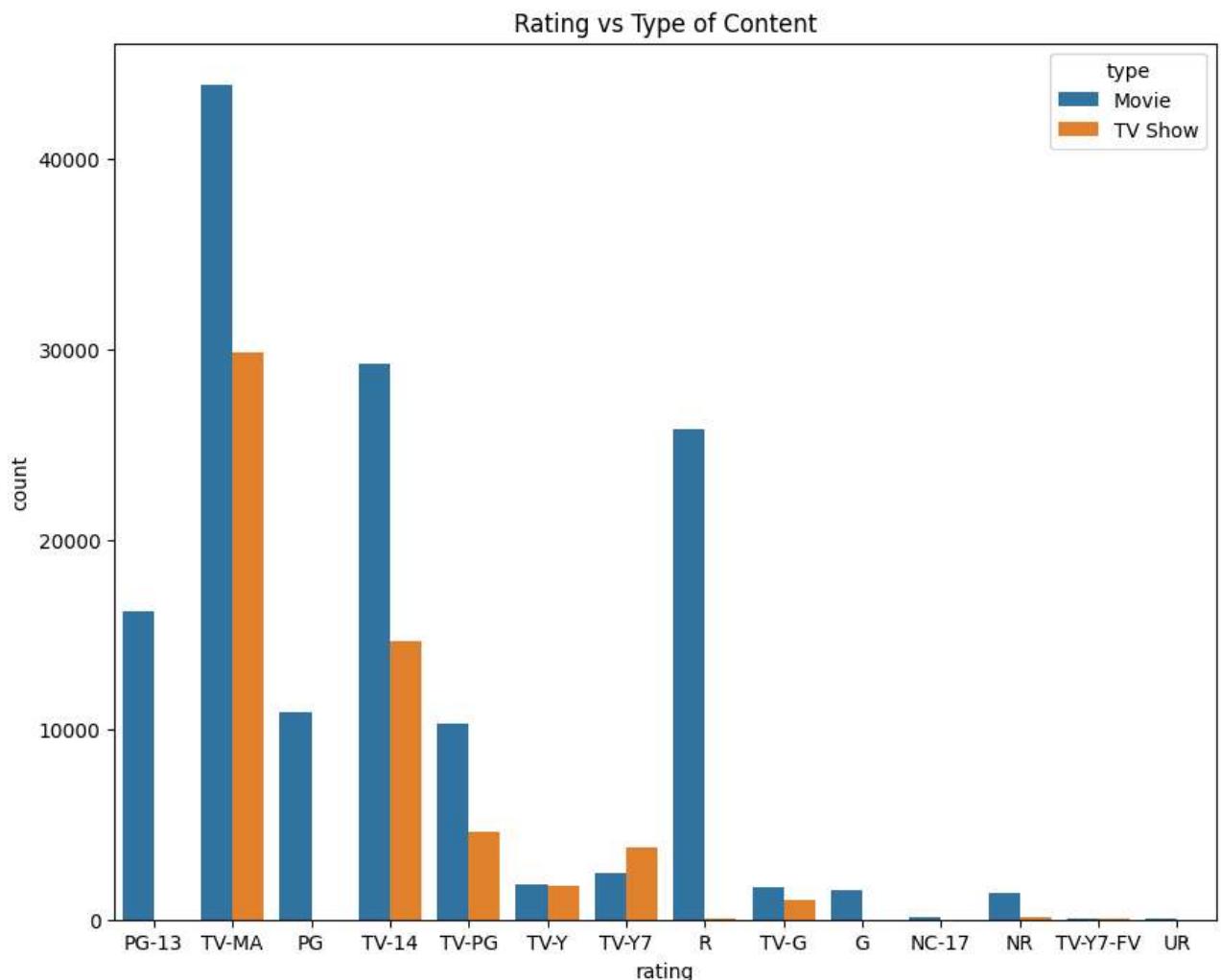
```
In [317]: 1 nf_du['release_year'].plot(kind='hist',bins=20)
2 plt.title("Frequency of movies released per year")
```

```
Out[317]: Text(0.5, 1.0, 'Frequency of movies released per year')
```

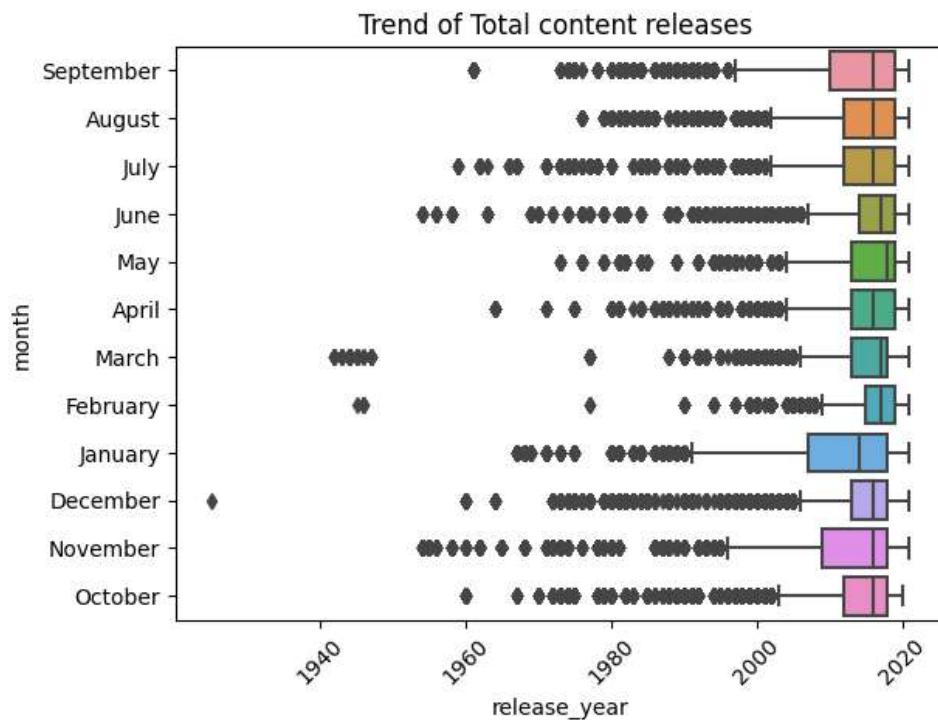


```
In [318]: 1 plt.figure(figsize=(10,8))
2 sns.countplot(x='rating', hue='type', data=nf_du)
3 plt.title("Rating vs Type of Content")
```

Out[318]: Text(0.5, 1.0, 'Rating vs Type of Content')

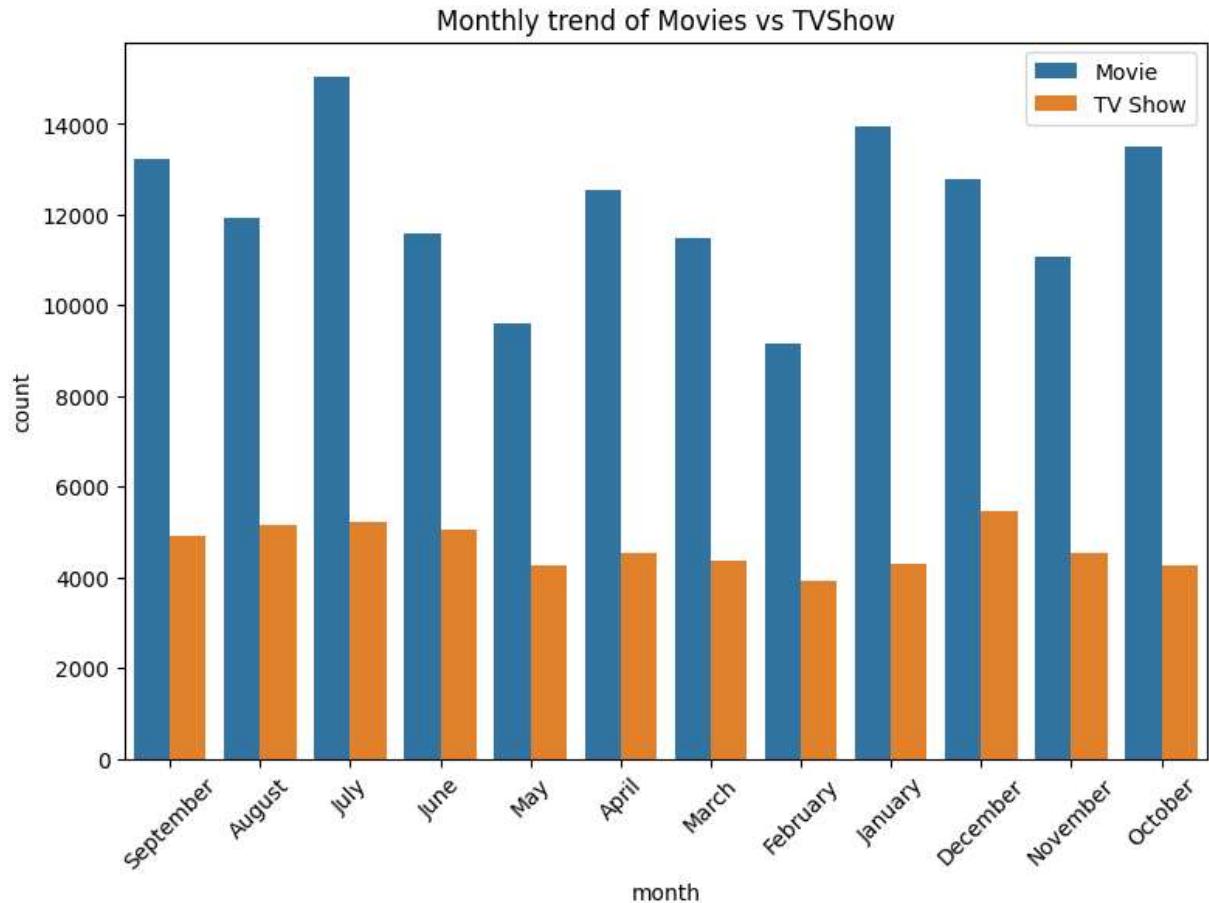


```
In [191]: 1 sns.boxplot(data=nf_du,y='month',x='release_year')
2 plt.xticks(rotation=45)
3 plt.tick_params(axis='x', labelsize=10)
4 plt.title("Trend of Total content releases ")
5 plt.show()
6
```



```
In [342]: 1 # m=nf_du.groupby(['month','type']).nunique().set_index('month')
2 plt.figure(figsize=(9,6))
3 sns.countplot(data=nf_du, x='month',hue='type')
4 plt.xticks(rotation=45)
5 plt.tick_params(axis='x', labelsize=10)
6 plt.legend(loc='upper right')
7 plt.title("Monthly trend of Movies vs TVShow")
```

Out[342]: Text(0.5, 1.0, 'Monthly trend of Movies vs TVShow')

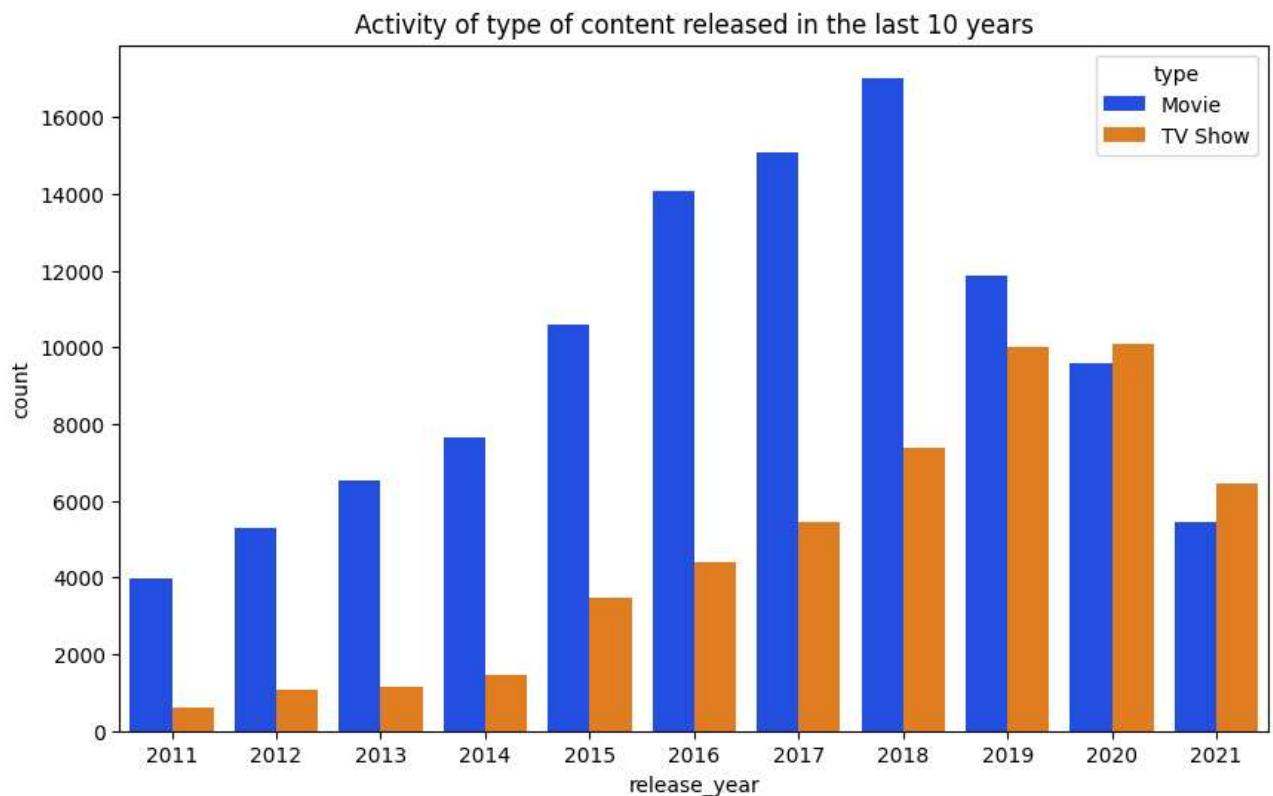


```
In [125]: 1 last10years= nf_du[['type', "release_year"]]
2 last10years = last10years.loc[last10years["release_year"]>=2011]
3 (last10years.shape[0]/nf_du.shape[0])*100
```

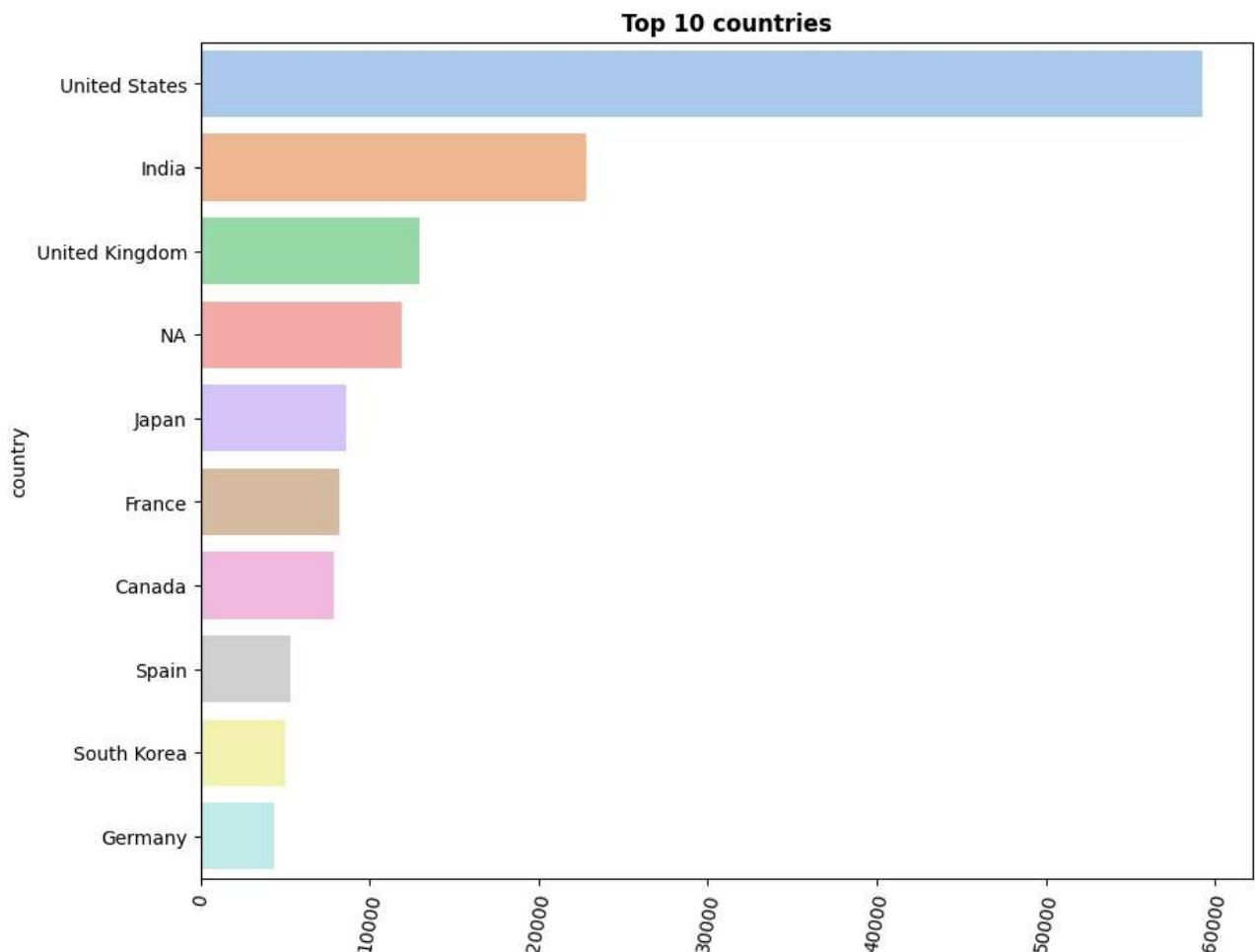
Out[125]: 78.61859718008209

```
In [344]: 1 last10years_df = last10years.groupby("release_year")["type"].size().reset_index()
2 last10years_df = pd.DataFrame(last10years_df)
3 last10years_df
4
5 plt.figure(figsize = (10,6))
6 count_plot = sns.countplot(x = "release_year", data = last10years, hue="type",
7                             palette= "bright")
8 count_plot.set(title = "Activity of type of content released in the last 10 years")
9
```

Out[344]: [Text(0.5, 1.0, 'Activity of type of content released in the last 10 years')]



```
In [348]: 1 plt.figure(figsize = (10,8))
2 country_plot = sns.barplot(x = nf_du.country.value_counts()[:10].values,
3                             y= nf_du.country.value_counts()[:10].index,palette = "pastel")
4 plt.xticks(rotation=80)
5 plt.title("Top 10 countries",fontweight="bold")
6 plt.show()
```



```
In [312]: 1
```

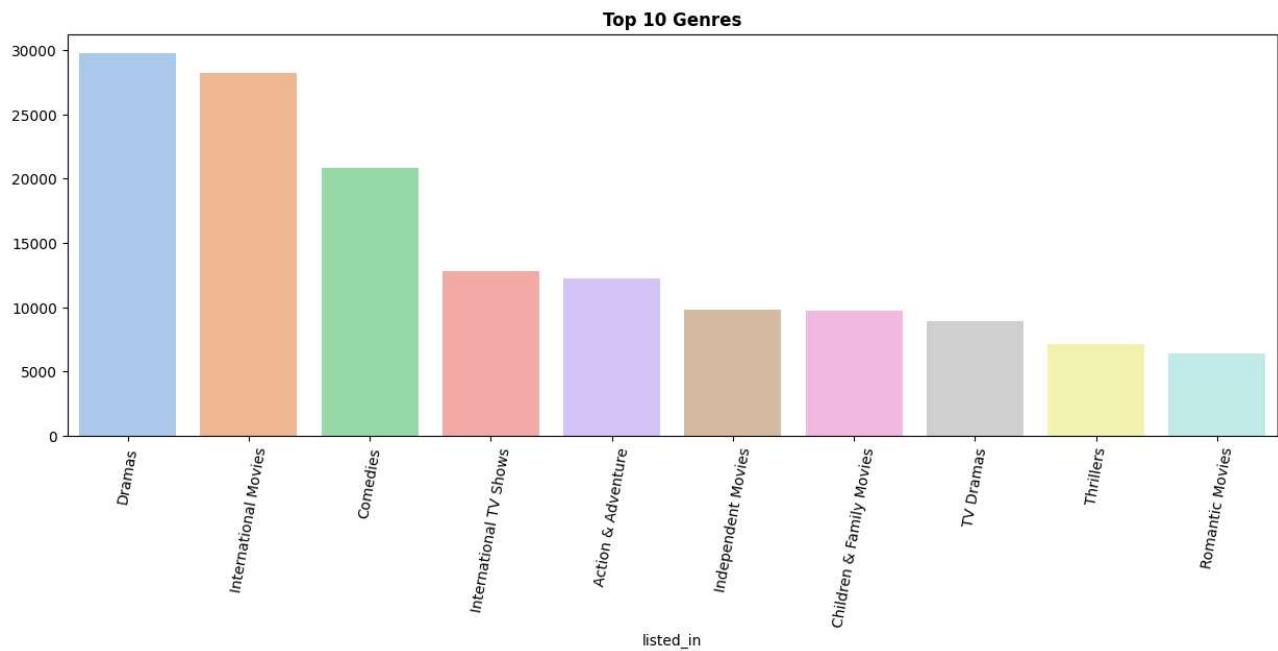
```
Out[312]: (42,)
```

In [346]:

```

1 plt.figure(figsize=(15,5))
2 sns.barplot(x = nf_du["listed_in"].value_counts().head(10).index,
3             y = nf_du["listed_in"].value_counts().head(10).values, palette="pastel")
4 plt.xticks(rotation=80)
5 plt.title("Top 10 Genres", fontweight="bold")
6 plt.show()

```



In [315]:

```

1 6.#Insights based on Non Graphical and Visual Analysis
2 #some observations on range of attributes and their relationship b/w them
3
4 #1.The dataset has about 8807 rows and 12 columns.
5
6 #2a.The 'released_year' attribute is the date on which the movies were released in theatres,some movies o
7 # on the streaming platform itself and data_added conatins date when movies that were released in theat
8 # so both attributes are very much inter-related
9
10 #2b.The 'type' attribute conatins either movies or TV shows and Listed_in attribute is a sub-category of
11 # It also related to 'duration' where movies have duration in minutes and TVShows have duration in mult
12
13 #3.univariate plot give us information related to either single categorical or single numerical
14 # the pie chart represents univariate categorical using type categorical variable,
15 # 2 histogram plots are represented using released_year where the distribution is skewed relative to the
16 # bivariate chart includes two variables for analysis. In this analysis countplot is used on rating and ty
17 # boxplot on month and release_year , two countplots on month vs type and type vs release_year,barplot o
18

```

```
In [ ]: 1 # Business Insights  
2  
3 # 1.The number of movie content on the streaming service is more than 2.5 times the TV Shows. This goes o  
4 # movies is still the the pouplar content on the paltform.  
5  
6 # 2.In the last 10 years trend chart we can see that the streaming service gained popularity and reached  
7 # offering in 2018 and then started seeing a downward trend, this can be attributed to password sharing  
8 # the covid pandemic surge that started from late 2019 to 2021  
9 # However the TVShows saw a jump in 2020 and 2021 making Long format content popular.  
10  
11 # 3. Even In boxplot chart we can see that the 75% of the total content released were between 2017-2021 :  
12 # the profitablity was acheived  
13  
14 # 4. The top 10 genre data suggest that Drama genre was the most popular and romantic movies were least j  
15  
16 # 5. Ratings chart suggests the featured movies having TV-MA rating were targeted at mature audience.  
17  
18 # 6. USA produces the most amount of content on Netflix with 33.6% whereas india stands second with unite  
19
```

```
In [ ]: 1 # Recommendations:  
2 # 1.The platform must focus on producing more Good Quality TV shows as people enjoy watching for Long ho  
3  
4 # 2.The platform also need to produce more content for non-english speaking regions as they can tap into  
5 # and stand to gain more subsrcibers apart from US,UK regions  
6
```

```
In [ ]: 1
```