Case Study on Netflix data

In [1]:
```
#Business Problem

#Analyze the data and generate insights that could help Netflix ijn deciding whic
#to produce and how they can grow the business in different countries
```

In [2]:
```
#Importing the dataset using pandas
import pandas as pd
NFDF = pd.read_csv(r"H:\Scaler\Pandas\Netflix Project Work\netflix.csv")
```

In [3]:
```
#Checking the first 5 rows of the dataset
NFDF.head()
```

Out[3]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	dura
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	Sea:
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	Se:
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	Se:
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	Sea:

In [4]: `#Checking the info in the dataset`
`NFDF.info()`

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

In [5]: `NFDF.shape`

Out[5]: `(8807, 12)`

In [6]: `#The Netflix dataset has 8807 rows and 12 columns. 11 of the 12 columns are of ob`
`#release_year is int64`
`#There are also null values present in dataset in the director,cast,country,durat`

In [7]: `NFDF.describe()`

Out[7]:

	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 [8]: `NFDF['type'].nunique()`

Out[8]: `2`

In [9]:
```python
NFDF['type'].unique()
```

Out[9]: array(['Movie', 'TV Show'], dtype=object)

In [10]:
```python
#There are only 2 categories under the unique column which are Movie and TV Show
#movies and TV Shows .Splitting the dataset into movies and TV Shows will help us
#A new dataframe called NFMV is being created with contains the type as only movi
#NFTV which contains only TV shows related information .
```

In [11]:
```python
NFMV = NFDF[NFDF['type']=='Movie']
NFMV.head()
```

Out[11]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG 1
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	NaN	September 24, 2021	2021	PG
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV MA
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG 1
12	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	2021	TV MA

In [12]:
```python
#creating a dataframe for TV Shows
NFTV = NFDF[NFDF['type']=='TV Show']
```

In [13]: `NFTV`

Out[13]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA
...	...	...	...	...	...	...	...	...	...
8795	s8796	TV Show	Yu-Gi-Oh! Arc-V	NaN	Mike Liscio, Emily Bauer, Billy Bob Thompson, ...	Japan, Canada	May 1, 2018	2015	TV-Y7
8796	s8797	TV Show	Yunus Emre	NaN	Gökhan Atalay, Payidar Tüfekçioglu, Baran Akbu...	Turkey	January 17, 2017	2016	TV-PG
8797	s8798	TV Show	Zak Storm	NaN	Michael Johnston, Jessica Gee-George, Christin...	United States, France, South Korea, Indonesia	September 13, 2018	2016	TV-Y7

	show_id	type	title	director	cast	country	date_added	release_year	rating
8800	s8801	TV Show	Zindagi Gulzar Hai	NaN	Sanam Saeed, Fawad Khan, Ayesha Omer, Mehreen ...	Pakistan	December 15, 2016	2012	TV-PG
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7

2676 rows × 12 columns

In [14]:
```python
#Checking the ratio between movies and tv shows in the dataset
round(len(NFMV)/len(NFDF),1)
```

Out[14]: 0.7

In [15]:
```python
round(len(NFTV)/len(NFDF),2)
```

Out[15]: 0.3

In [16]:
```python
#70 percent of the data is of movies and 30 percent of the data belongs to the TV
```

In [17]:
```python
NFDF['rating'].value_counts()
```

Out[17]:
```
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
```

In [18]:
```python
#TV-MA has the most count of movies and tv shows from the entire data set .
```

```
In [19]: NFDF.groupby(['type'])['rating'].value_counts()
```

```
Out[19]: type      rating
         Movie     TV-MA        2062
                   TV-14        1427
                   R             797
                   TV-PG         540
                   PG-13         490
                   PG            287
                   TV-Y7         139
                   TV-Y          131
                   TV-G          126
                   NR             75
                   G              41
                   TV-Y7-FV        5
                   NC-17           3
                   UR              3
                   66 min          1
                   74 min          1
                   84 min          1
         TV Show   TV-MA        1145
                   TV-14         733
                   TV-PG         323
                   TV-Y7         195
                   TV-Y          176
                   TV-G           94
                   NR              5
                   R               2
                   TV-Y7-FV        1
         Name: rating, dtype: int64
```

```
In [20]: #Upon checking the ratings for the movies and tv shows independantly the rating c
         #movies and TV shows .
```

```
In [21]: #Checking the average run time of movies from the NFMV dataframe
```

```
In [22]: NFMV['duration'] = NFMV['duration'].apply(lambda x: float(str(x).split(' ')[0]))
```

```
C:\Users\india\AppData\Local\Temp\ipykernel_22452\2827027116.py:1: SettingWithC
opyWarning:
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/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pyd
ata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c
opy)
  NFMV['duration'] = NFMV['duration'].apply(lambda x: float(str(x).split(' ')
[0]))
```

```
In [23]: round(NFMV['duration'].mean())
```

```
Out[23]: 100
```

In [24]: `#The mean runtime of movies is 100 minutes .`

In [25]: `NFMV.describe()`

Out[25]:

	release_year	duration
count	6131.000000	6128.000000
mean	2013.121514	99.577187
std	9.678169	28.290593
min	1942.000000	3.000000
25%	2012.000000	87.000000
50%	2016.000000	98.000000
75%	2018.000000	114.000000
max	2021.000000	312.000000

In [26]: `#The minimum runtime of movies is 3 minutes and  the maximum runtime is 312 minut`

In [27]: `#Checking the minimum , maximum and average duration for TV shows`

In [28]: `NFTV['duration'] = NFTV['duration'].apply(lambda x: float(str(x).split(' ')[0]))`

```
C:\Users\india\AppData\Local\Temp\ipykernel_22452\2887938341.py:1: SettingWithC
opyWarning:
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/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pyd
ata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c
opy)
  NFTV['duration'] = NFTV['duration'].apply(lambda x: float(str(x).split(' ')
[0]))
```

In [29]: `NFTV.head()`

Out[29]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	dur
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	

In [30]: `round(NFTV['duration'].mean())`

Out[30]: 2

In [31]: `#The average number of seasons for TV Shows seems to be 2.`

In [32]: `NFTV['duration'].max()`

Out[32]: 17.0

In [33]: `NFTV['duration'].min()`

Out[33]: 1.0

In [34]: `#The maximum number of seasons for TV show seems to be 17 and the minimum number`

In [35]: `NFTV[NFTV['duration']==17.0]`

Out[35]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	durat
548	s549	TV Show	Grey's Anatomy	NaN	Ellen Pompeo, Sandra Oh, Katherine Heigl, Just...	United States	July 3, 2021	2020	TV-14	1

In [36]: `#The Longest running TV show with 17 seasons was Grey's Anatomy .It was from the`
`#Romantic TV Shows and TV Dramas`

In [37]: `NFMV[NFMV['duration']==312]`

Out[37]:

|	show_id	type	title	director	cast	country	date_added	release_year	ratin
4253	s4254	Movie	Black Mirror: Bandersnatch	NaN	Fionn Whitehead, Will Poulter, Craig Parkinson...	United States	December 28, 2018	2018	TV M/

In [38]: `#The movie with the longest runtime is Black mirror : Bandersnatch which had a ro`
`# to the genre of Dramas,International movies ,Sci-Fi and Fantasy.`

In [39]: `NFTV['duration'].value_counts()`

Out[39]:
```
1.0      1793
2.0       425
3.0       199
4.0        95
5.0        65
6.0        33
7.0        23
8.0        17
9.0         9
10.0        7
13.0        3
15.0        2
12.0        2
11.0        2
17.0        1
Name: duration, dtype: int64
```

In [40]: `#The TV shows with just 1 season seems to be of the highest count .The number of`

In [41]:
```python
import matplotlib.pyplot as plt
NFTV['duration'].value_counts().plot(kind='bar')
plt.xlabel('Number of seasons')
plt.ylabel('Count of series')
plt.title('Number of seasons per TV series')
plt.show()
```
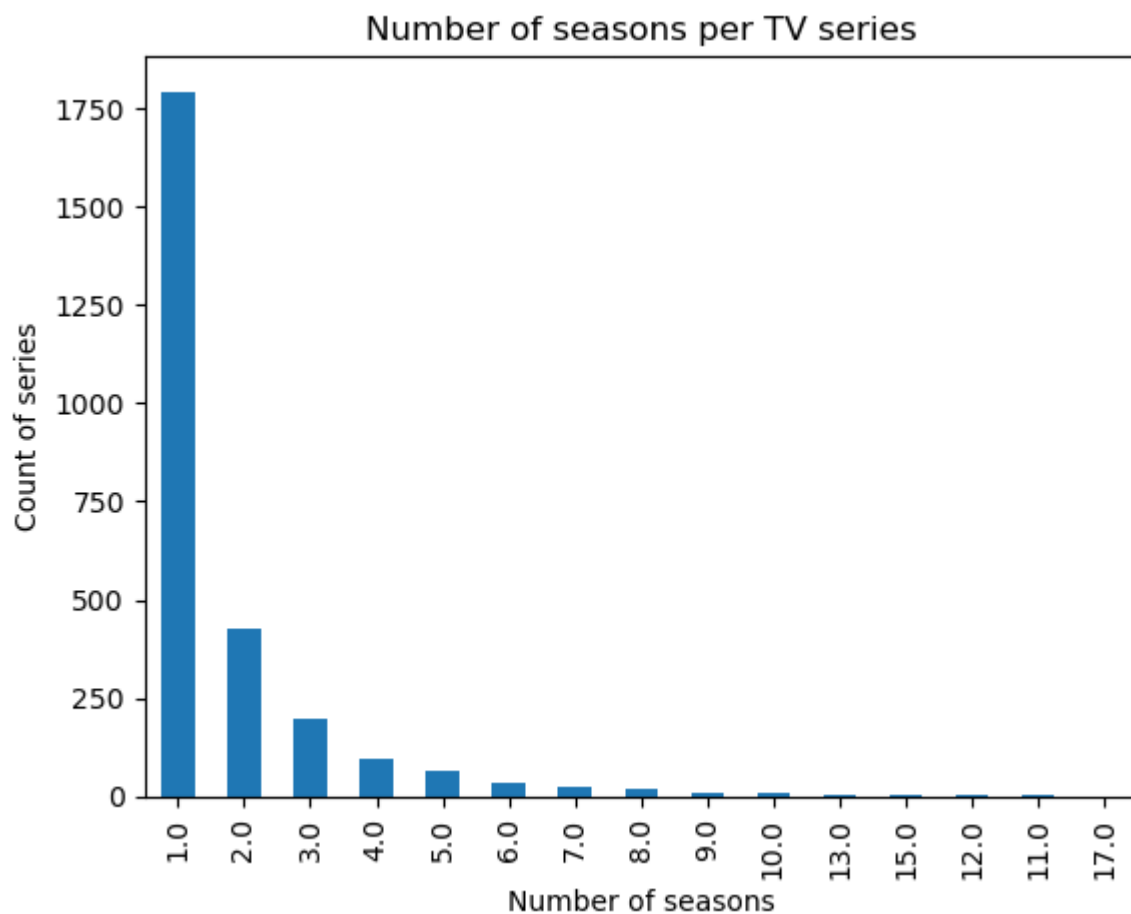

In [42]:
```python
#The column release_year can be converted to datetime data type
```

In [180]:
```python
NFDF['release_year'] = pd.to_datetime(NFDF['release_year'])
```

In [181]:
```python
NFDF.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 14 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   datetime64[ns]
 7   release_year 8797 non-null   datetime64[ns]
 8   rating       8803 non-null   object
 9   duration     8804 non-null   float64
 10  listed_in    8807 non-null   object
 11  description  8807 non-null   object
 12  Month_Added  8797 non-null   float64
 13  Year_Added   8797 non-null   float64
dtypes: datetime64[ns](2), float64(3), object(9)
memory usage: 963.4+ KB
```

In [45]:
```python
#Similarly the date_added column also has year, month and date information .
#It can also be converted to datetime data type
```

In [182]:
```python
NFDF['date_added'] = pd.to_datetime(NFDF['date_added'])
```

In [183]:
```python
NFDF.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 14 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   datetime64[ns]
 7   release_year 8797 non-null   datetime64[ns]
 8   rating       8803 non-null   object
 9   duration     8804 non-null   float64
 10  listed_in    8807 non-null   object
 11  description  8807 non-null   object
 12  Month_Added  8797 non-null   float64
 13  Year_Added   8797 non-null   float64
dtypes: datetime64[ns](2), float64(3), object(9)
memory usage: 963.4+ KB
```

In [48]: *# further adding columns for month and year of the title being added to netflix*

In [184]:
```python
NFDF['Month_Added'] = NFDF['date_added'].dt.month
NFDF['Year_Added'] = NFDF['date_added'].dt.year
```

In [185]: NFDF

Out[185]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2021-09-25	P
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021-09-24	T M
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	2021-09-24	2021-09-24	T M
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	2021-09-24	2021-09-24	T M
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021-09-24	2021-09-24	T M
...	...	...	...	...	...	...	...	...	
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	2019-11-20	2019-11-20	
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2019-07-01	TV-
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	2019-11-01	2019-11-01	

	show_id	type	title	director	cast	country	date_added	release_year	rati
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	2020-01-11	2020-01-11	F
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	2019-03-02	2019-03-02	TV-

8807 rows × 14 columns

In [186]: `NFDF.groupby('Year_Added')['title'].count()`

Out[186]:
```
Year_Added
2008.0       2
2009.0       2
2010.0       1
2011.0      13
2012.0       3
2013.0      11
2014.0      24
2015.0      82
2016.0     429
2017.0    1188
2018.0    1649
2019.0    2016
2020.0    1879
2021.0    1498
Name: title, dtype: int64
```

In [52]: `NFDF['Year_Added'].min()`

Out[52]: `2008.0`

In [53]: `NFDF['Year_Added'].max()`

Out[53]: `2021.0`

In [54]: *#In the given dataset titles have been added to netflix from the year 2008 .As pe*
*#added per year is 2 and it is in the year 2008 and 2009 and the maximum number o*
*#total of 2016 titles being added .*

In [55]:
```python
NFDF.groupby('Year_Added')['title'].count().plot(kind='bar')
plt.xlabel('Year of addition')
plt.ylabel('Count of titles added')
plt.title('Number of titles added per year')
plt.plot()
```

Out[55]: []
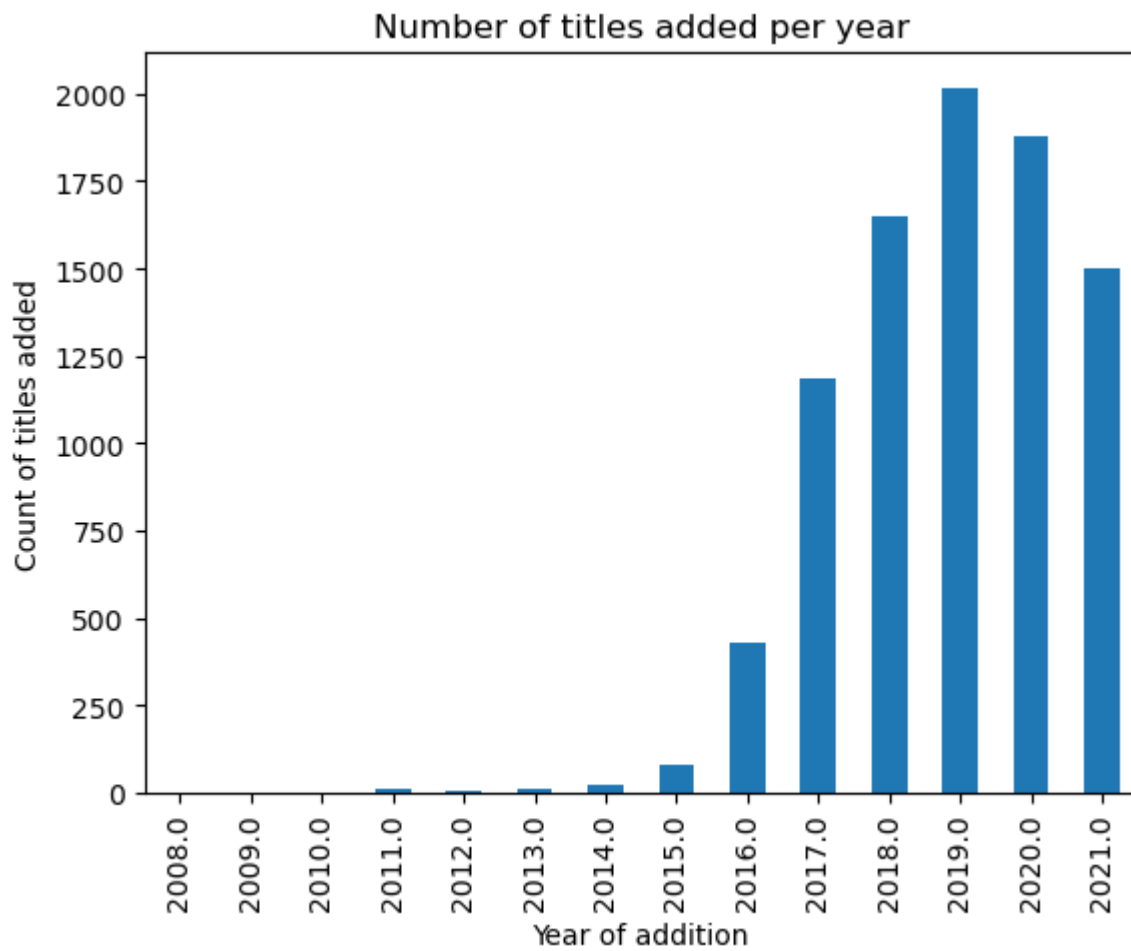
In [56]:
```python
NFDF['duration'] = NFDF['duration'].apply(lambda x: float(str(x).split(' ')[0]))
```

In [57]:
```python
def split_values(df,C):
    constraint = df[C].apply(lambda x: str(x).split(', ')).tolist()
    NFDF_new = pd.DataFrame(constraint, index = df['title'])
    NFDF_new = NFDF_new.stack()
    NFDF_new = pd.DataFrame(NFDF_new)
    NFDF_new.reset_index(inplace= True)
    NFDF_new = NFDF_new[['title',0]]
    NFDF_new.columns = ['title',C]
    df = df.merge(NFDF_new, left_on = 'title', right_on='title')
    df.drop([C + '_x'],axis =1 , inplace = True)
    df[C] = df[C + '_y']
    df.drop([C+'_y'],axis=1,inplace=True)
    return df
```

In [58]:
```python
#Unnesting the data in the cast column using the split_values function
```

In [59]:
```python
NFDFC = split_values(NFDF,"cast")
```

In [61]: NFDFC

Out[61]:

	show_id	type	title	director	country	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	2021-09-25	2021-09-25	PG-13	90.0
1	s2	TV Show	Blood & Water	NaN	South Africa	2021-09-24	2021-09-24	TV-MA	2.0
2	s2	TV Show	Blood & Water	NaN	South Africa	2021-09-24	2021-09-24	TV-MA	2.0
3	s2	TV Show	Blood & Water	NaN	South Africa	2021-09-24	2021-09-24	TV-MA	2.0
4	s2	TV Show	Blood & Water	NaN	South Africa	2021-09-24	2021-09-24	TV-MA	2.0
...	...	...	...	...	...	...	...	...	...
64946	s8807	Movie	Zubaan	Mozez Singh	India	2019-03-02	2019-03-02	TV-14	111.0
64947	s8807	Movie	Zubaan	Mozez Singh	India	2019-03-02	2019-03-02	TV-14	111.0
64948	s8807	Movie	Zubaan	Mozez Singh	India	2019-03-02	2019-03-02	TV-14	111.0
64949	s8807	Movie	Zubaan	Mozez Singh	India	2019-03-02	2019-03-02	TV-14	111.0

	show_id	type	title	director	country	date_added	release_year	rating	duration
64950	s8807	Movie	Zubaan	Mozez Singh	India	2019-03-02	2019-03-02	TV-14	111.0

64951 rows × 14 columns

In [ ]: *#After unnesting the cast column the count of rows increases to 64951*

In [ ]: *#Performing similar unnesting on the listed_in ,director and country columns*

In [ ]: *#Unnesting the director column*

In [62]:
```python
NFDFCD = split_values(NFDFC,"director")
```

In [63]: NFDFCD

Out[63]:

	show_id	type	title	country	date_added	release_year	rating	duration	liste
0	s1	Movie	Dick Johnson Is Dead	United States	2021-09-25	2021-09-25	PG-13	90.0	Document
1	s2	TV Show	Blood & Water	South Africa	2021-09-24	2021-09-24	TV-MA	2.0	Interna TV Show Drama Mys
2	s2	TV Show	Blood & Water	South Africa	2021-09-24	2021-09-24	TV-MA	2.0	Interna TV Show Drama Mys
3	s2	TV Show	Blood & Water	South Africa	2021-09-24	2021-09-24	TV-MA	2.0	Interna TV Show Drama Mys
4	s2	TV Show	Blood & Water	South Africa	2021-09-24	2021-09-24	TV-MA	2.0	Interna TV Show Drama Mys
...	...	...	...	...	...	...	...	...	
751259	s8807	Movie	Zubaan	India	2019-03-02	2019-03-02	TV-14	111.0	Dra Interna Movies, N & Mus
751260	s8807	Movie	Zubaan	India	2019-03-02	2019-03-02	TV-14	111.0	Dra Interna Movies, N & Mus
751261	s8807	Movie	Zubaan	India	2019-03-02	2019-03-02	TV-14	111.0	Dra Interna Movies, N & Mus
751262	s8807	Movie	Zubaan	India	2019-03-02	2019-03-02	TV-14	111.0	Dra Interna Movies, N & Mus

	show_id	type	title	country	date_added	release_year	rating	duration	liste
751263	s8807	Movie	Zubaan	India	2019-03-02	2019-03-02	TV-14	111.0	Dra Interna Movies, & Mu:

751264 rows × 14 columns

In [65]: *#Unnesting the listed_in column from the original data frame*

In [64]: `NFDFG = split_values(NFDF,"listed_in")`

In [67]: `NFDFG.head()`

Out[67]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	dura
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2021-09-25	PG-13	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021-09-24	TV-MA	
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021-09-24	TV-MA	
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021-09-24	TV-MA	
4	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	2021-09-24	2021-09-24	TV-MA	

In [68]: *#Unnesting the countries column from the newly created NFDFG*
`NFDFCG = split_values(NFDFG,"cast")`

In [69]: NFDFCG

Out[69]:

	show_id	type	title	director	cast	date_added	release_year	rating	duration	d
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	2021-09-25	2021-09-25	PG-13	90.0	fa
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0	
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0	
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0	
4	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0	
...	...	...	...	...	...	...	...	...	...	
58563	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2019-03-02	TV-14	111.0	h
58564	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2019-03-02	TV-14	111.0	h
58565	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2019-03-02	TV-14	111.0	h

	show_id	type	title	director	cast	date_added	release_year	rating	duration	d
58566	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2019-03-02	TV-14	111.0	h
58567	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2019-03-02	TV-14	111.0	h

58568 rows × 14 columns

In [70]: `NFDFCG.head()`

Out[70]:

	show_id	type	title	director	cast	date_added	release_year	rating	duration	descri
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	2021-09-25	2021-09-25	PG-13	90.0	A father the e h fil
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0	cro path p; Cape
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0	cro path p; Cape
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0	cro path p; Cape
4	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0	cro path p; Cape

In [71]: `NFDFCD.head()`

Out[71]:

	show_id	type	title	country	date_added	release_year	rating	duration	listed_in	de
0	s1	Movie	Dick Johnson Is Dead	United States	2021-09-25	2021-09-25	PG-13	90.0	Documentaries	fa
1	s2	TV Show	Blood & Water	South Africa	2021-09-24	2021-09-24	TV-MA	2.0	International TV Shows, TV Dramas, TV Mysteries	C
2	s2	TV Show	Blood & Water	South Africa	2021-09-24	2021-09-24	TV-MA	2.0	International TV Shows, TV Dramas, TV Mysteries	C
3	s2	TV Show	Blood & Water	South Africa	2021-09-24	2021-09-24	TV-MA	2.0	International TV Shows, TV Dramas, TV Mysteries	C
4	s2	TV Show	Blood & Water	South Africa	2021-09-24	2021-09-24	TV-MA	2.0	International TV Shows, TV Dramas, TV Mysteries	C

In [189]: `NFDFCD.groupby(['director','cast'])['cast'].count()`

Out[189]:
```
director        cast
A. L. Vijay     G.V. Prakash Kumar      6
                Hema                    8
                Joy Mathew              8
                Munishkanth             6
                Murli Sharma            8
                                       ..
Şenol Sönmez    Seda Güven              8
                Somer Karvan            8
                Yosi Mizrahi            8
                Zerrin Sümer            8
                Özgür Emre Yıldırım     8
Name: cast, Length: 62741, dtype: int64
```

In [85]: `NFDFCG.groupby(['listed_in','country'])['listed_in'].count().sort_values()`

Out[85]:
```
listed_in                 country
Children & Family Movies  Kazakhstan        1
Documentaries             United States,    1
                          Hong Kong         1
Stand-Up Comedy           Iceland           1
                          Netherlands       1
                                          ...
Independent Movies        United States  1020
Comedies                  United States  1492
Dramas                    India          1848
                          United States  1902
International Movies       India          2373
Name: listed_in, Length: 1464, dtype: int64
```

In [86]: `NFDFCG.groupby(['country','listed_in'])['country'].count().sort_values()`

Out[86]:
```
country        listed_in
Poland         Documentaries                1
Argentina      Movies                       1
Kenya          Documentaries                1
Kazakhstan     Children & Family Movies     1
Singapore      Stand-Up Comedy              1
                                          ...
United States  Independent Movies        1020
               Comedies                  1492
India          Dramas                    1848
United States  Dramas                    1902
India          International Movies       2373
Name: country, Length: 1464, dtype: int64
```

In [94]:
```python
NFDFG['listed_in'].value_counts().head(20).plot(kind='bar')
plt.xlabel('Genre')
plt.ylabel('Count of titles')
plt.title('Distribution of titles by genres')
```

Out[94]: Text(0.5, 1.0, 'Distribution of titles by genres')
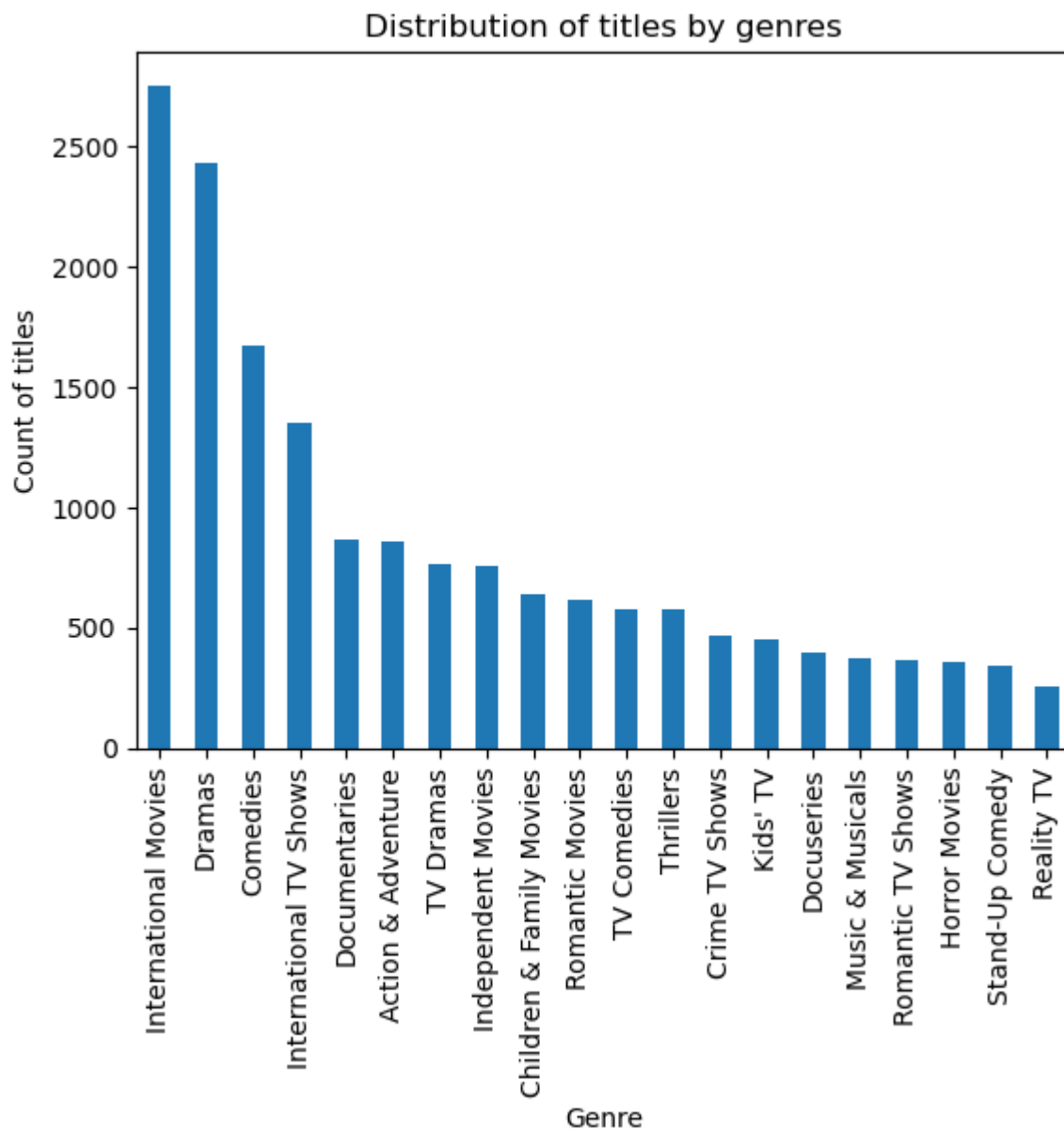
In [92]: `#International movies and TV shows,dramas and comedies seem to be the most viewed`

In [190]: `NFDFG['listed_in'].value_counts().head(20)`

Out[190]:
```
International Movies         2752
Dramas                      2427
Comedies                    1674
International TV Shows       1351
Documentaries                869
Action & Adventure           859
TV Dramas                    763
Independent Movies           756
Children & Family Movies     641
Romantic Movies              616
TV Comedies                  581
Thrillers                    577
Crime TV Shows               470
Kids' TV                     451
Docuseries                   395
Music & Musicals             375
Romantic TV Shows            370
Horror Movies                357
Stand-Up Comedy              343
Reality TV                   255
Name: listed_in, dtype: int64
```

In [105]: `NFDF.min()`

```
C:\Users\india\AppData\Local\Temp\ipykernel_22452\26993687.py:1: FutureWarning:
Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None')
is deprecated; in a future version this will raise TypeError.  Select only vali
d columns before calling the reduction.
  NFDF.min()
```

Out[105]:
```
show_id                                              s1
type                                              Movie
title                                            #Alive
date_added                          2008-01-01 00:00:00
release_year                        2008-01-01 00:00:00
duration                                            1.0
listed_in                            Action & Adventure
description      "Bridgerton" cast members share behind-the-sce...
Month_Added                                         1.0
Year_Added                                       2008.0
dtype: object
```

In [106]: `import seaborn as sns`

In [107]: `sns.countplot(x='type',data=NFDF)`

Out[107]: `<AxesSubplot:xlabel='type', ylabel='count'>`
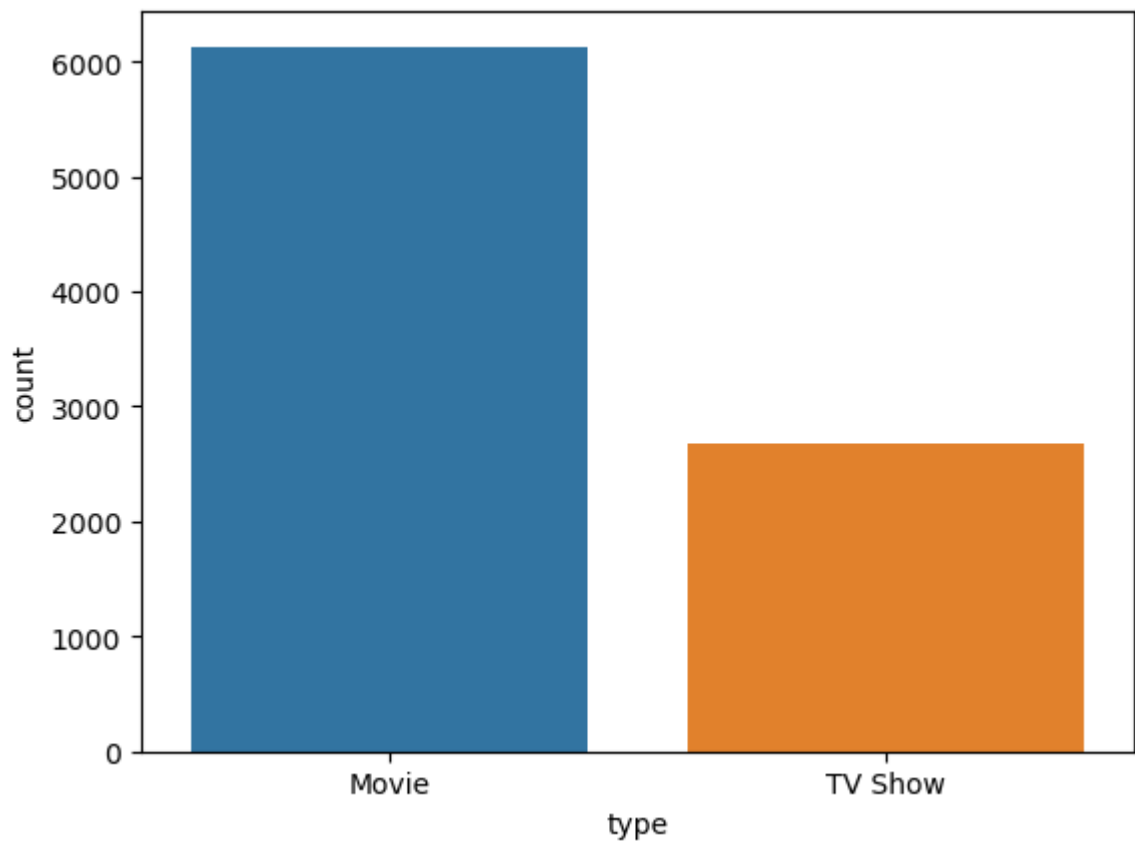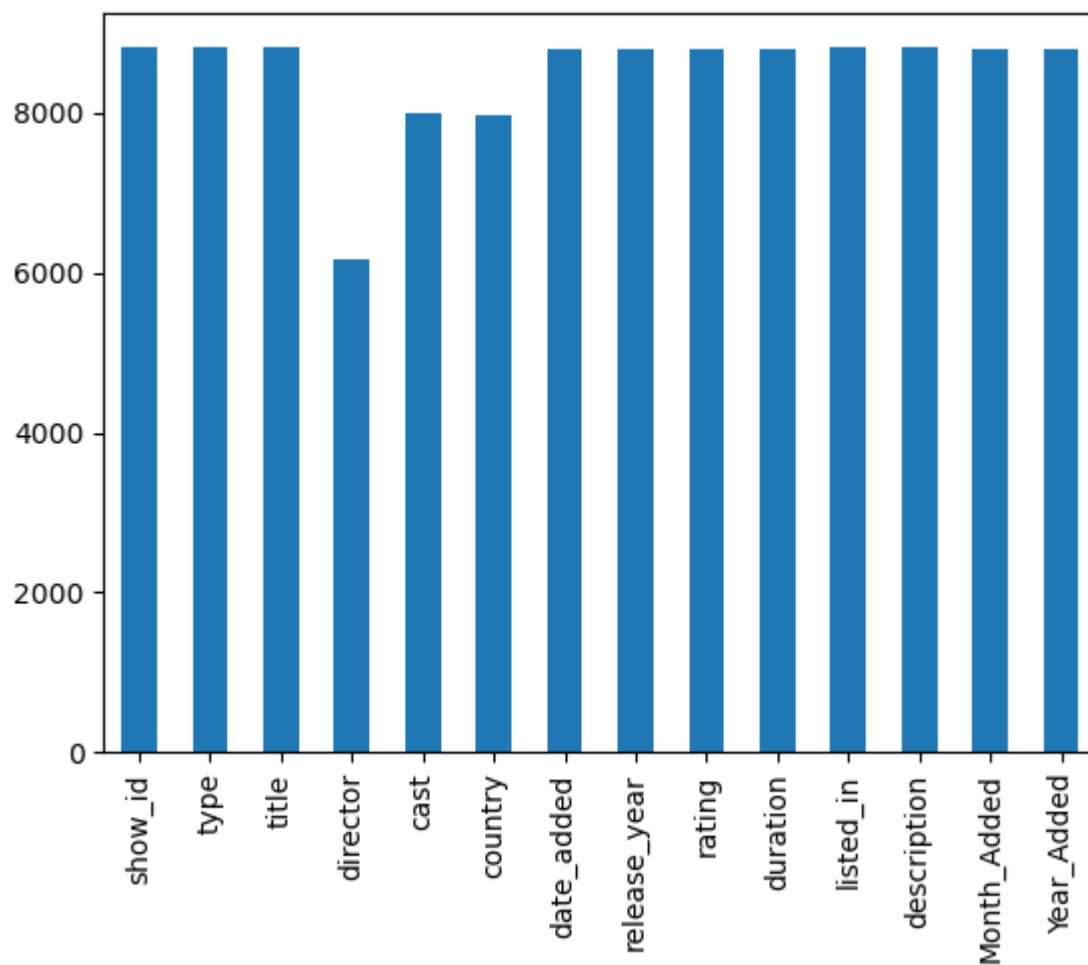
In [112]: `sns.boxplot(x='type',y='Month_Added',data=NFDFCG,palette='rainbow')`
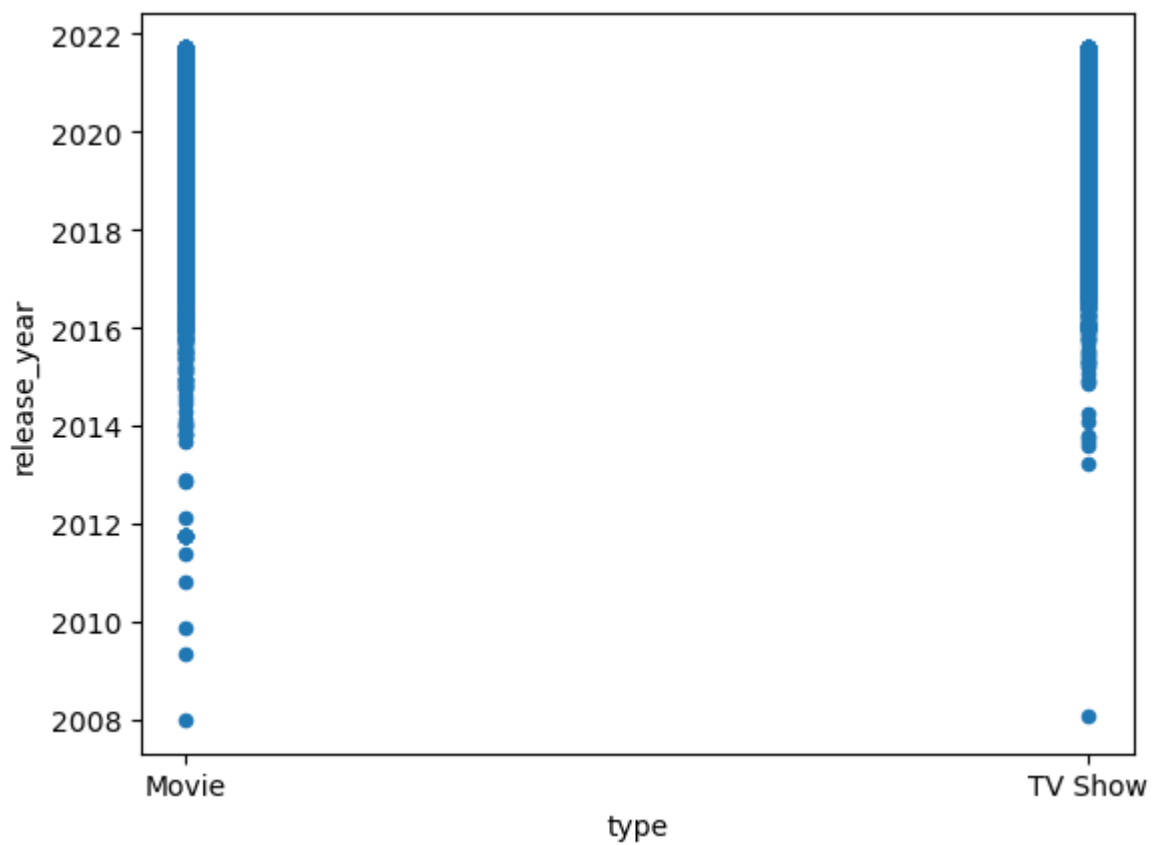
Out[112]: `<AxesSubplot:xlabel='type', ylabel='Month_Added'>`
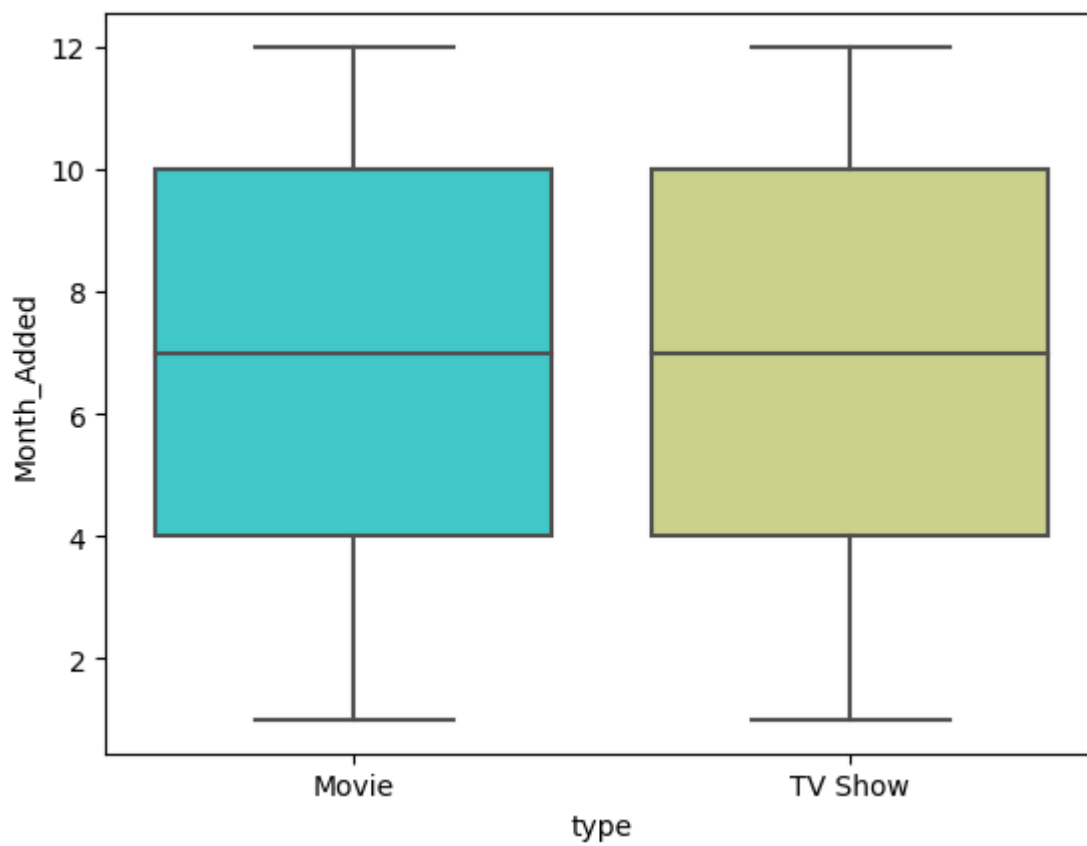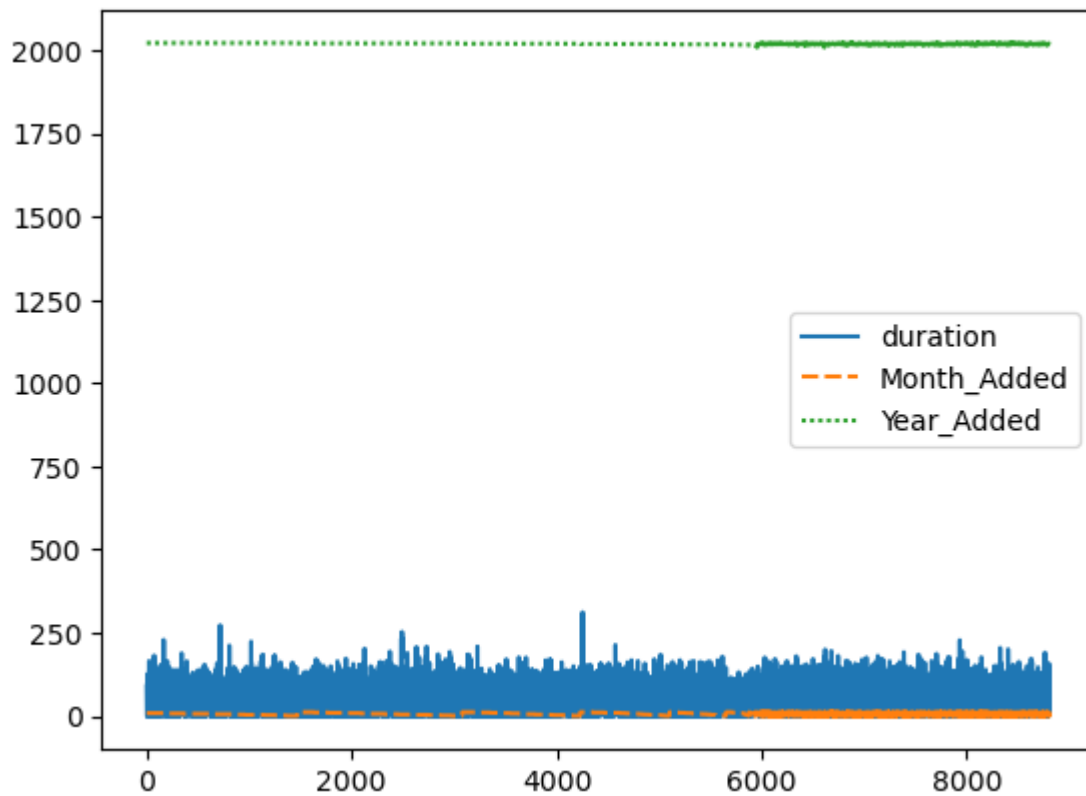
In [126]:
```
NFDF.count().plot.bar()
```

Out[126]: `<AxesSubplot:>`

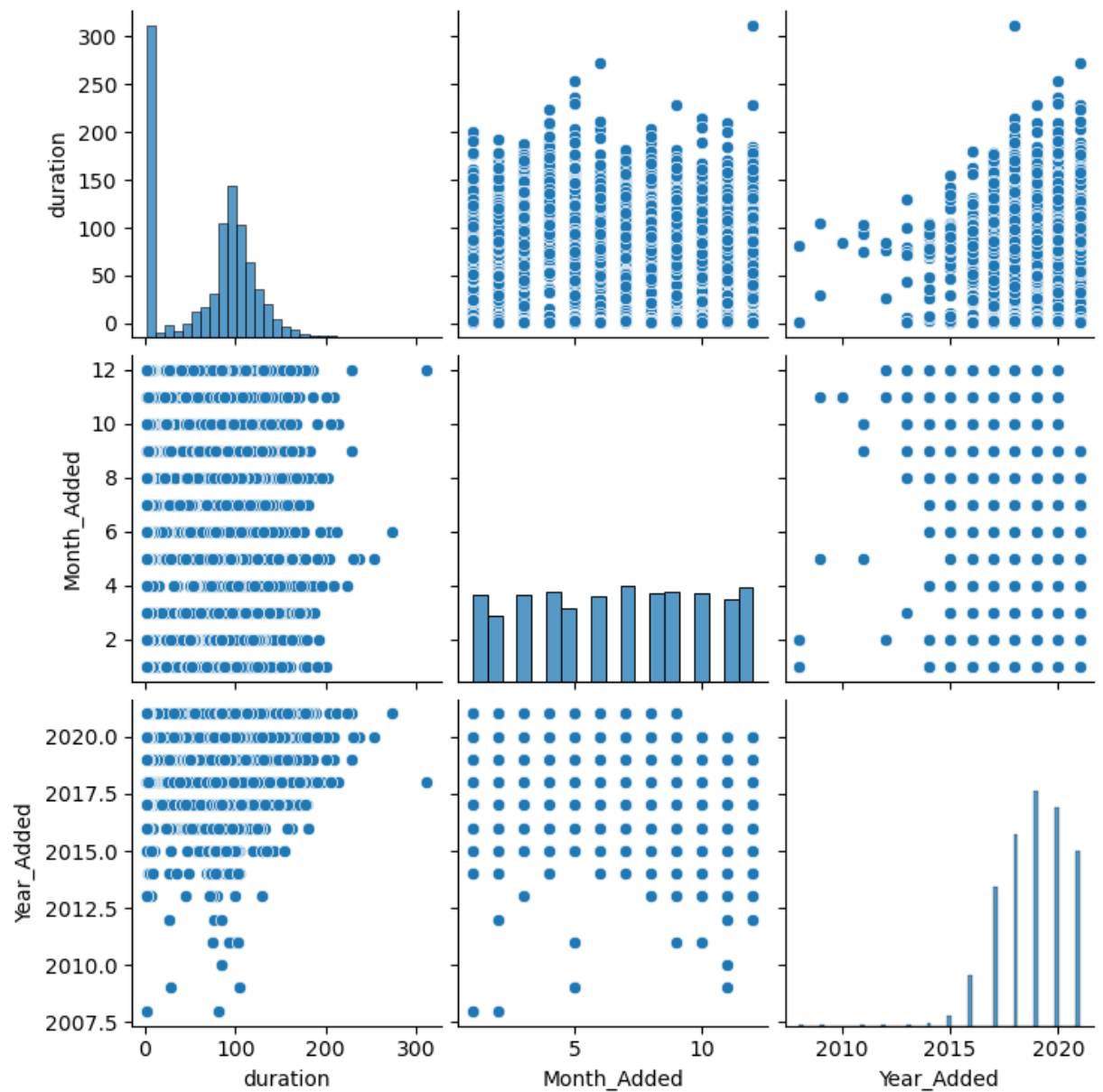In [131]: `NFDF.plot.scatter(x='type',y='release_year')`
`plt.show()`
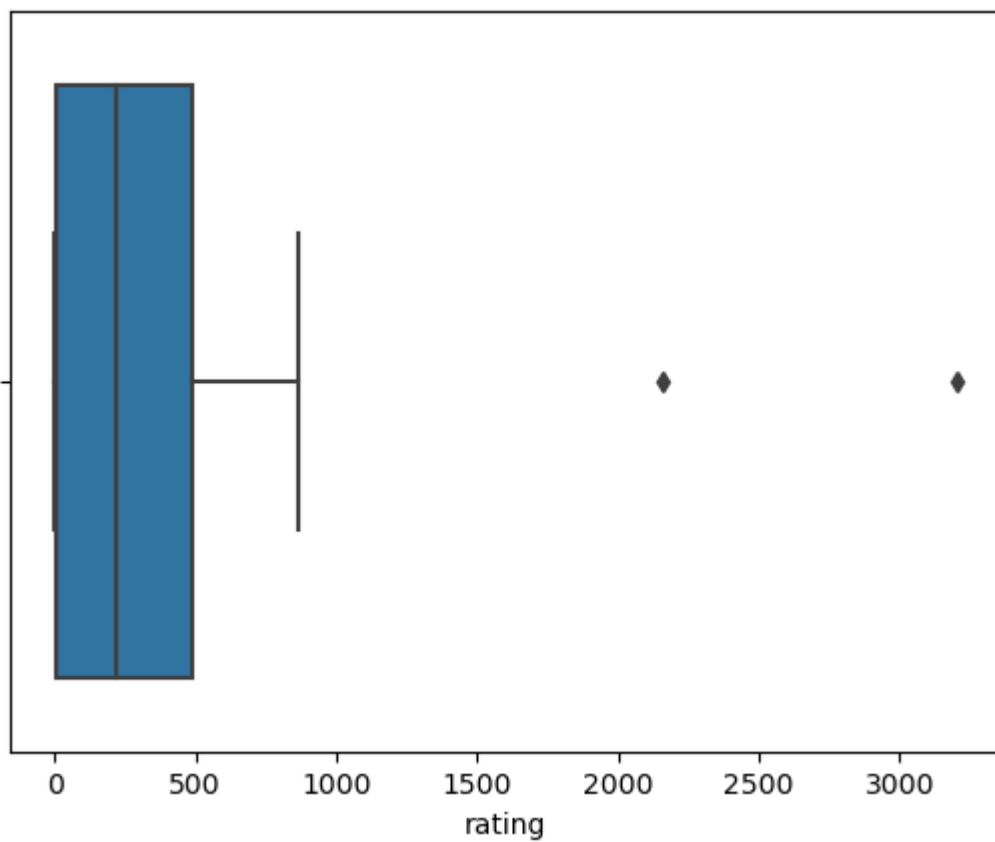
In [140]:
```python
sns.lineplot(data=NFDF)
plt.show()
```

In [141]: `sns.pairplot(data=NFDF)`

Out[141]: `<seaborn.axisgrid.PairGrid at 0x18970e0dd60>`

In [143]:
```python
sns.boxplot(data = NFDF
            ,x = NFDF['rating'].value_counts()
            )
```

Out[143]: <AxesSubplot:xlabel='rating'>

In [146]:
```python
NFDF['rating'].value_counts()
```

Out[146]:
```
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
```

In [157]: `NFDF.groupby(['listed_in','country'])['country'].count().head(10).plot(kind='bar'`

Out[157]: `<AxesSubplot:xlabel='listed_in,country'>`
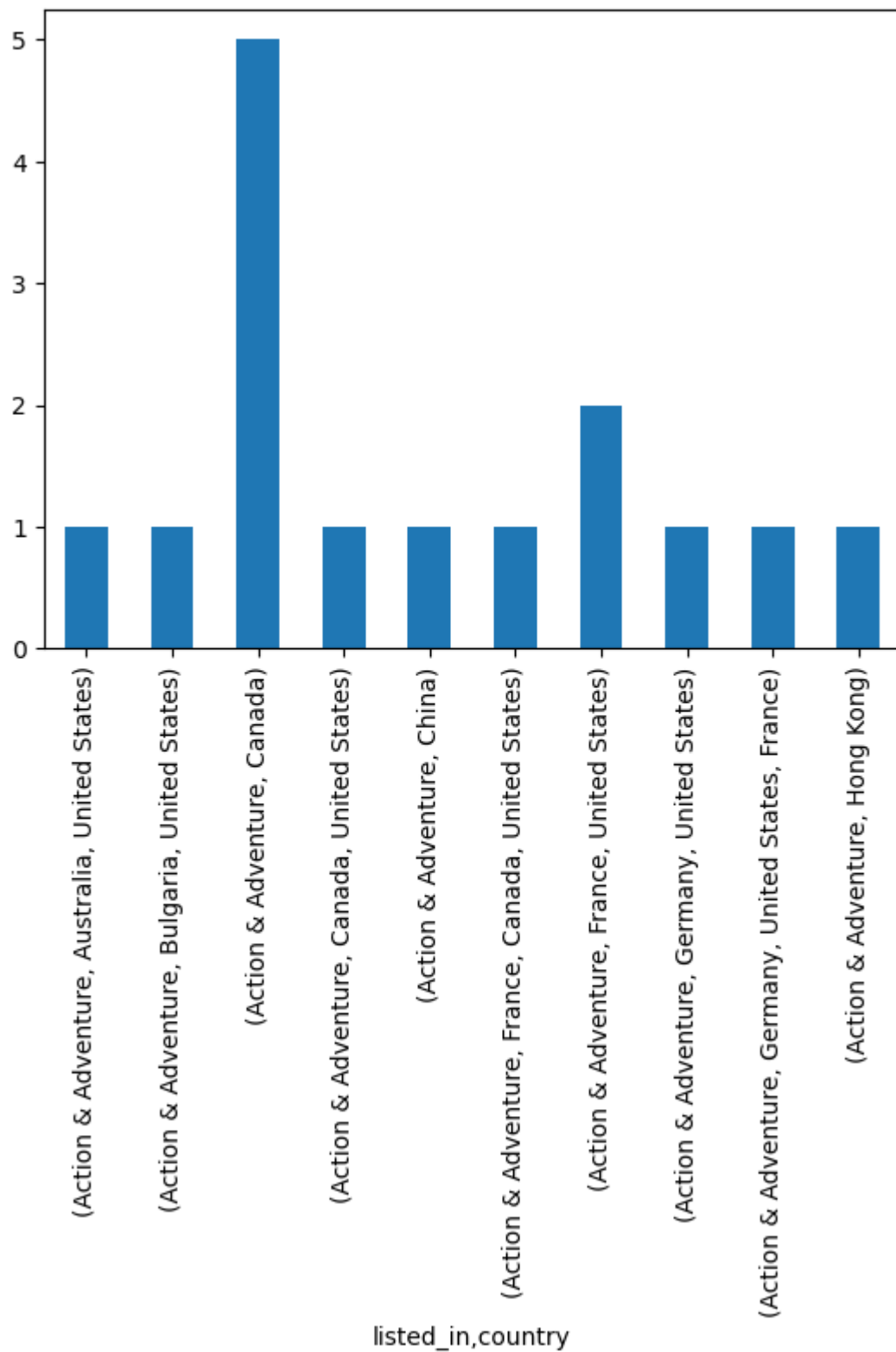
In [162]: `sns.heatmap(NFDF.corr(), cmap="YlGnBu", annot=True);`
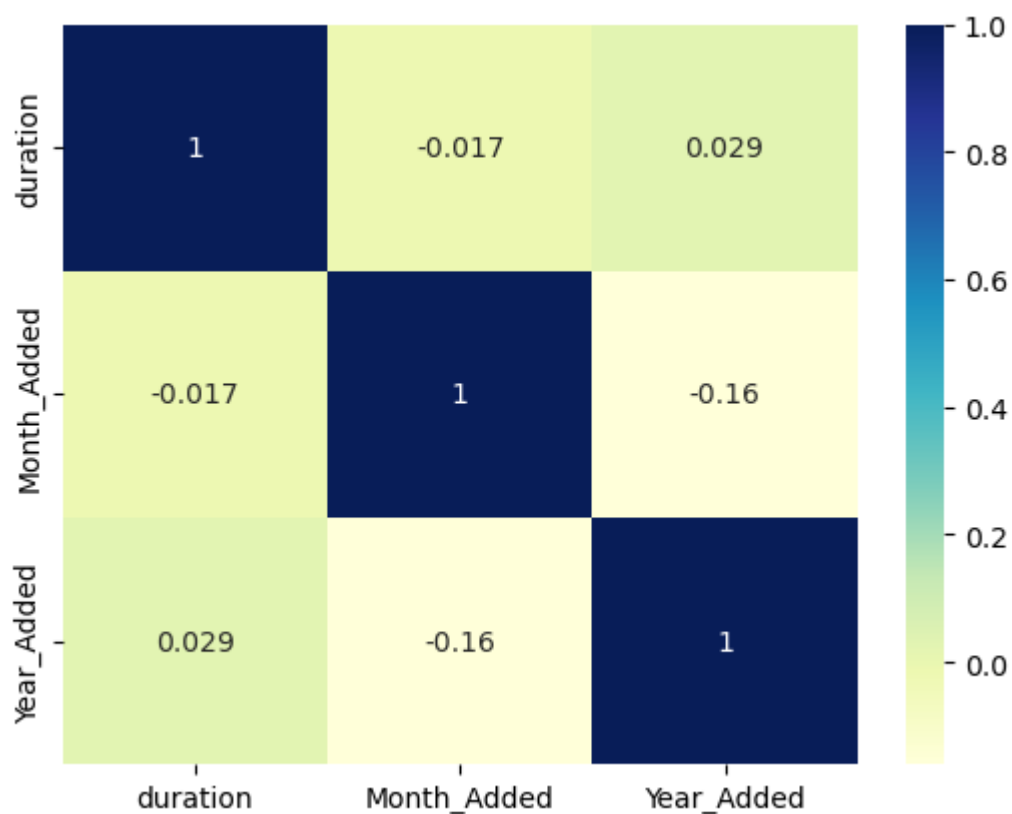
In [167]: `NFMV['director'].value_counts()`

Out[167]:
```
Rajiv Chilaka                  19
Raúl Campos, Jan Suter         18
Suhas Kadav                    16
Marcus Raboy                   15
Jay Karas                      14
                               ..
Dennis Rovira van Boekholt      1
Naoto Amazutsumi                1
Jenny Gage                      1
Kaila York                      1
Mozez Singh                     1
Name: director, Length: 4354, dtype: int64
```

```
In [168]:  NFTV['director'].value_counts()
```

```
Out[168]:  Alastair Fothergill                          3
           Rob Seidenglanz                             2
           Hsu Fu-chun                                 2
           Iginio Straffi                              2
           Shin Won-ho                                 2
                                                      ..
           Juliana Vicente                             1
           Chang Chin-jung, Chen Rong-hui             1
           Thierry Demaizière, Alban Teurlai          1
           Manolo Caro                                 1
           Michael Cumming                             1
           Name: director, Length: 222, dtype: int64
```

```
In [177]:  NFDF['cast'].value_counts().head().plot(kind='pie',explode=[0.1,0.2,0.3,0.4,0.5])
```
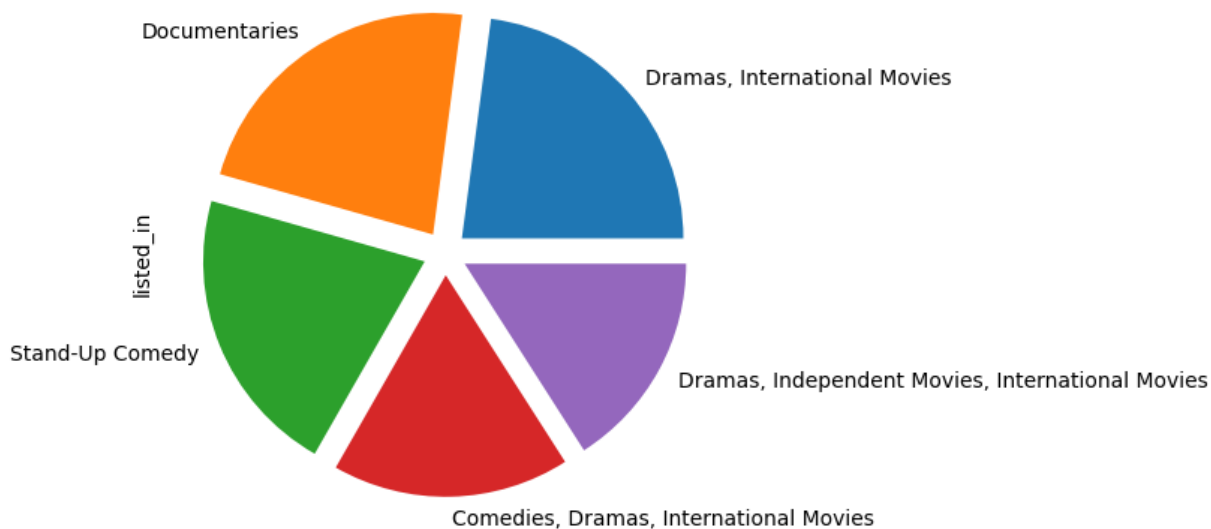
```
Out[177]:  <AxesSubplot:ylabel='cast'>
```
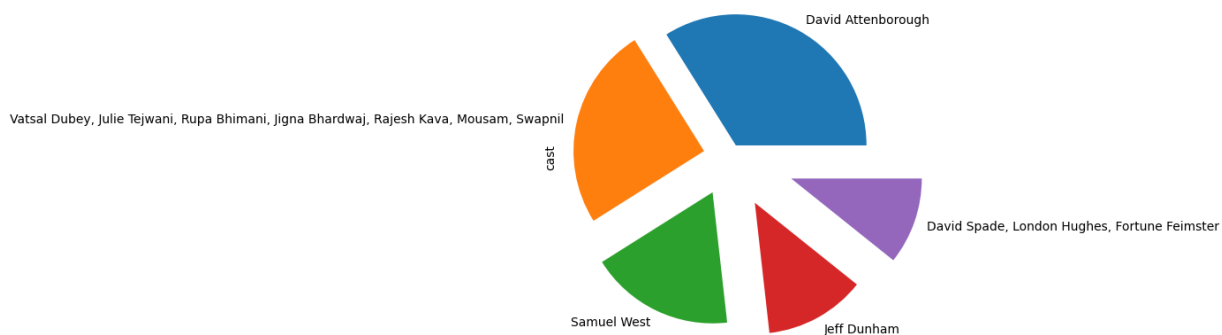


```
In [179]:  NFDF['listed_in'].value_counts().head().plot(kind='pie',explode=[0.1,0.1,0.1,0.1,
```

```
Out[179]:  <AxesSubplot:ylabel='listed_in'>
```

In [191]: `NFDF['listed_in'].value_counts()`

Out[191]:
```
Dramas, International Movies                          362
Documentaries                                        359
Stand-Up Comedy                                      334
Comedies, Dramas, International Movies               274
Dramas, Independent Movies, International Movies     252
                                                     ...
Kids' TV, TV Action & Adventure, TV Dramas             1
TV Comedies, TV Dramas, TV Horror                      1
Children & Family Movies, Comedies, LGBTQ Movies       1
Kids' TV, Spanish-Language TV Shows, Teen TV Shows     1
Cult Movies, Dramas, Thrillers                         1
Name: listed_in, Length: 514, dtype: int64
```

In [192]: `NFDF['cast'].value_counts()`

Out[192]:
```
David Attenborough
19
Vatsal Dubey, Julie Tejwani, 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 Ko
bayashi, Ken Yasuda, Arata Furuta, Suzuki Matsuo, Koichi Yamadera, Arata Iura,
Chikako Kaku, Kotaro Yoshida        1
Toyin Abraham, Sambasa Nzeribe, Chioma Chukwuka Akpotha, Chioma Omeruah, Chiwet
alu Agu, Dele Odule, Femi Adebayo, Bayray McNwizu, Biodun Stephen
1
Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary, Meghna Mali
k, Malkeet Rauni, Anita Shabdish, Chittaranjan Tripathy
1
Name: cast, Length: 7692, dtype: int64
```

In [193]: `NFDFI = NFDF[NFDF['country']=='India']`

In [194]: `NFDFI['rating'].value_counts()`

Out[194]:
```
TV-14        550
TV-MA        248
TV-PG        134
TV-Y7         14
TV-G           9
TV-Y           5
PG-13          4
NR             4
PG             2
R              1
TV-Y7-FV       1
Name: rating, dtype: int64
```

In [195]: `NFDFUS = NFDF[NFDF['country']=='United States']`

In [196]: `NFDFUS['rating'].value_counts()`

Out[196]:
```
TV-MA        928
R            440
TV-14        406
PG-13        289
TV-PG        251
PG           164
TV-Y7        101
TV-Y          87
TV-G          82
NR            35
G             29
84 min         1
66 min         1
74 min         1
TV-Y7-FV       1
NC-17          1
UR             1
Name: rating, dtype: int64
```

In [204]: NFDFCG

Out[204]:

	show_id	type	title	director	cast	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	2021-09-25	2021-09-25	PG-13	90.0
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0
4	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021-09-24	TV-MA	2.0
...	...	...	...	...	...	...	...	...	...
58563	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2019-03-02	TV-14	111.0
58564	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2019-03-02	TV-14	111.0
58565	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2019-03-02	TV-14	111.0

	show_id	type	title	director	cast	date_added	release_year	rating	duration
58566	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2019-03-02	TV-14	111.0
58567	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2019-03-02	TV-14	111.0

58568 rows × 14 columns

In [214]: `NFDFCD['cast'].value_counts()`

Out[214]:
```
nan                 960
Kristen Bell        738
Liev Schreiber      641
Terrence Howard     634
Elizabeth Banks     629
                   ...
Vince Carter          1
Anthony Lister        1
Jonathan Kydd         1
Roberto Palazuelos    1
Nick Broomfield       1
Name: cast, Length: 36440, dtype: int64
```

In [222]: `NFDFJC = NFDFCG[NFDFCG['country']=='Japan']`

In [225]: `NFDFJC['listed_in'].value_counts()`

Out[225]:
```
International TV Shows         354
Anime Series                  318
International Movies          193
Anime Features               170
Action & Adventure           158
Romantic TV Shows             61
TV Dramas                     54
Dramas                        54
Kids' TV                      49
Crime TV Shows                46
Children & Family Movies      45
Teen TV Shows                 39
TV Comedies                   26
Sci-Fi & Fantasy              24
Reality TV                    20
Comedies                      20
Independent Movies            19
TV Thrillers                  18
Romantic Movies               17
TV Horror                     15
Documentaries                 15
Music & Musicals              15
TV Action & Adventure         14
Thrillers                     13
TV Mysteries                  12
Horror Movies                 10
Classic Movies                 9
Cult Movies                    3
British TV Shows               3
Docuseries                     3
LGBTQ Movies                   3
Movies                         2
Stand-Up Comedy & Talk Shows   2
Sports Movies                  2
TV Shows                       1
Name: listed_in, dtype: int64
```

In [226]: `NFDFJC['rating'].value_counts()`

Out[226]:
```
TV-MA    631
TV-14    599
TV-PG    296
TV-Y7     95
PG        56
PG-13     50
R         42
TV-Y      17
G          9
TV-G       8
Name: rating, dtype: int64
```

In [ ]: