

Double-click (or enter) to edit

1.Defining the problem statement

To analyze the data and generate insights that could help Netflix in deciding which type of shows/movies to produce and how they can grow the business in different countries

2.To analyze basic metrics, we have to get the data and load it into a dataframe

```
from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
import io
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
path = "/content/drive/MyDrive/Business_case_datasets/netflix_dataset.csv"
df = pd.read_csv(path)
df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas,	NaN	September 24, 2021	2021	TV-MA	1 Season

```
len(df)

8807
```

```
#Getting the type of data in the dataframe
df.info()
```

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

```
df.shape

(8807, 12)
```

```
#getting list of all columns present in the df
df.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object', length=12)
```

```
dtype='object')
```

```
df.nunique()
```

```
show_id      8807
type          2
title        8807
director     4528
cast         7692
country       748
date_added   1767
release_year   74
rating        17
duration     220
listed_in     514
description   8775
dtype: int64
```

```
#checking how many null values are present in each column in the data
```

```
df.isnull().sum()
```

```
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
```

```
df['rating'].value_counts()
```

```
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
```

```
#getting count of movies in each year sorted in desc order
```

```
df['release_year'].value_counts(sort=True)
```

```
2018      1147
2017      1032
2019      1030
2020       953
2016       902
...
1959        1
1925        1
1961        1
1947        1
1966        1
Name: release_year, Length: 74, dtype: int64
```

```
df.head()
```

```
df['country'].value_counts()

United States    2818
India            972
United Kingdom   419
Japan            245
South Korea      199
...
Romania, Bulgaria, Hungary    1
Uruguay, Guatemala            1
France, Senegal, Belgium      1
Mexico, United States, Spain, Colombia    1
United Arab Emirates, Jordan    1
Name: country, Length: 748, dtype: int64

df_movie = df[df['type']=="Movie"]
df_movie.head()
```

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
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	91 min	Children & Family Movies	Equestria divided. But a bright-eyed hero be
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	125 min	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s

```
df_movie.index = np.arange(1,df_movie.shape[0]+1)
df_movie.head()
# df_movie.reset_index(drop=True)
```

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	
1	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...
2	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	91 min	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...
3	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s...

```
df_movie.nunique()

show_id    6131
type        1
title      6131
director   4354
cast       5445
country     651
date_added 1533
release_year    73
rating         17
```

```
duration      205
listed_in     278
description    6105
dtype: int64
```

```
df_movie['country'].value_counts()
```

```
United States      2058
India              893
United Kingdom     206
Canada            122
Spain              97
...
United Kingdom, Russia, United States    1
Paraguay, Argentina                     1
United Kingdom, Malawi                   1
Austria, Iraq, United States             1
United Arab Emirates, Jordan             1
Name: country, Length: 651, dtype: int64
```

```
df_show = df[df['type']=="TV Show"]
df_show.head(20)
```

show_id	type		title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mababane, Thabane...	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...

```
df_show.index = np.arange(1,df_show.shape[0]+1)
df_show.head(20)
```

```
5          s6      TV Show      Midnight Mass      Mike Flanagan      Gilford, Hamish      NaN      September 24, 2021      2021      TV-MA      1 Season      TV Horror, a charismatic young pris
```

```
df_show.nunique()
```

```
show_id      2676
type          1
title        2676
director      222
cast          2283
country       196
date_added   1051
release_year   46
rating        9
duration      15
listed_in     236
description   2672
dtype: int64
```

```
df_show['country'].value_counts()
```

```
United States      760
United Kingdom     213
Japan               169
South Korea         158
India               79
...
Belarus             1
United Kingdom, Australia      1
France, Australia, Germany     1
Australia, New Zealand, United States      1
United States, France, South Korea, Indonesia      1
Name: country, Length: 196, dtype: int64
```

```
df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas,	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV	To protect his family from a powerful drug

```
df_movie.nunique()
```

```
show_id      6131
type          1
title        6131
director     4354
cast         5445
country       651
date_added   1533
release_year   73
rating        17
duration      205
listed_in     278
description   6105
dtype: int64
```

We can see that number of movies produced are greater than number of TV shows produced

```
df_movie['rating'].value_counts()
```

```
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
74 min     1
84 min     1
66 min     1
Name: rating, dtype: int64
```

We can say that movies of type TV_MA rating are highest in number implying people prefer such kind of movies

```
#Let's manipulate movies df first and see NaN values,fill those and do some analysis
df_movie.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
1	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...
2	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	91 min	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...
3	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s...

```
#gives movies having nan in duration
df_movie[df_movie['duration'].isnull()]
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
3563	s5542	Movie	Louis C.K. 2017	Louis C.K.	Louis C.K.	United States	April 4, 2017	2017	74 min	NaN	Movies	Louis C.K. muses on religion, eternal love, gi...
3739	s5795	Movie	Louis C.K.: Hilarious	Louis C.K.	Louis C.K.	United States	September 16, 2016	2010	84 min	NaN	Movies	Emmy-winning comedy writer Louis C.K. brings h...
3748	s5814	Movie	Louis C.K.: Live at the Comedy Store	Louis C.K.	Louis C.K.	United States	August 15, 2016	2015	66 min	NaN	Movies	The comic puts his trademark hilarious/thought...

#dropping these rows as duration is nan and rating is corrupted

```
df_movie.dropna(subset=['duration'], how='all', inplace=True)
```

/usr/local/lib/python3.8/dist-packages/pandas/util/_decorators.py:311: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

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

```
return func(*args, **kwargs)
```

```
df_movie.isnull().sum()
```

```
show_id      0
type         0
title        0
director    188
cast        475
country     440
date_added   0
release_year 0
rating       2
duration     0
listed_in    0
description  0
dtype: int64
```

```
df_movie[df_movie['duration'].isnull()]
```

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
---------	------	-------	----------	------	---------	------------	--------------	--------	----------	-----------	-------------

```
df_movie['duration'] = df_movie['duration'].str.replace(r'min', '')

<ipython-input-362-87ef1dfdd974>:1: 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
df_movie['duration'] = df_movie['duration'].str.replace(r'min', '')
```

df_movie.head()

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90	Documentaries	As her father nears the end of his life, filmm...
2	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	91	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...
3	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s...

```
df_movie['duration'] = df_movie['duration'].astype(float)

<ipython-input-364-ebae8e7b55c8>:1: 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
df_movie['duration'] = df_movie['duration'].astype(float)
```

```
#getting average values of movie duration
df_movie['duration'].mean()

99.57718668407311

#Now filling NaN values with NOT AVAILABLE
# df_movie.fillna("NOT AVAILABLE",inplace=True)-----Did not run this

df_movie.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90.0	Documentaries	As her father nears the end of his life, filmm...
2	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	91.0	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...
3	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125.0	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s...

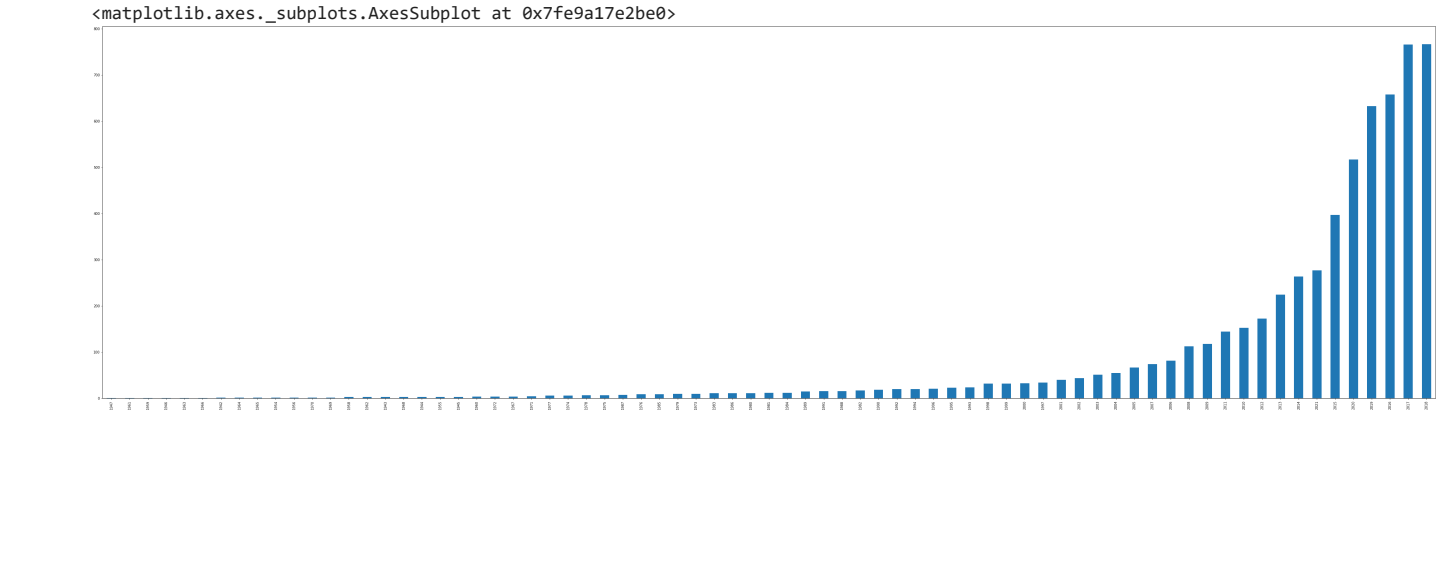
```
df_movie['release_year'].value_counts()
#Plot line chart or bar chart

2018    767
2017    766
```



```
2016    658
2019    633
2020    517
...
1966     1
1961     1
1946     1
1963     1
1947     1
Name: release_year, Length: 73, dtype: int64

df_movie[['release_year']].release_year.value_counts(ascending=True).plot(kind='bar',figsize=(70,20))
#we can see that the number of movies per year has increased over the years due to advancement in technology and increase in purchasing p
#thus this trend will continue to grow unless something unfortunate happens eg global pandemic,global economic slowdown,cyber attack etc
```



Unnesting the df for listed_in,cast,country and director

```
col1 = df_movie['director'].apply(lambda x:str(x).split(',')).tolist()
dfm1 = pd.DataFrame(col1,index=df_movie['title'])
dfm1 = dfm1.stack()
dfm1 = pd.DataFrame(dfm1.reset_index())
dfm1.rename(columns={0:'director'},inplace=True)
dfm1.drop(columns='level_1',inplace=True)
dfm1.head()
#gives title of movie and names of all directors who directed a particular movie
```

	title	director
0	Dick Johnson Is Dead	Kirsten Johnson
1	My Little Pony: A New Generation	Robert Cullen
2	My Little Pony: A New Generation	José Luis Ucha
3	Sankofa	Haile Gerima
4	The Starling	Theodore Melfi

```
dfm1['director'].value_counts()

nan          188
Rajiv Chilaka    22
Raúl Campos     18
Jan Suter       18
Suhas Kadav     16
...
Jumpei Mizusaki    1
Koji Morimoto     1
Michael Arias     1
Arthell Isom      1
Mozes Singh       1
Name: director, Length: 4886, dtype: int64
```

```
col2 = df_movie['cast'].apply(lambda x:str(x).split(',')).tolist()
col2
dfm2 = pd.DataFrame(col2,index=df_movie['title'])
dfm2 = dfm2.stack()
dfm2 = pd.DataFrame(dfm2.reset_index())
```

```
dfm2.rename(columns={0: 'actor'},inplace=True)
dfm2.drop(columns='level_1',inplace=True)
dfm2.head(15)
#gives title of movie and names of all actors who acted in a particular movie
```

	title	actor
0	Dick Johnson Is Dead	nan
1	My Little Pony: A New Generation	Vanessa Hudgens
2	My Little Pony: A New Generation	Kimiko Glenn
3	My Little Pony: A New Generation	James Marsden
4	My Little Pony: A New Generation	Sofia Carson
5	My Little Pony: A New Generation	Liza Koshy
6	My Little Pony: A New Generation	Ken Jeong
7	My Little Pony: A New Generation	Elizabeth Perkins
8	My Little Pony: A New Generation	Jane Krakowski
9	My Little Pony: A New Generation	Michael McKean
10	My Little Pony: A New Generation	Phil LaMarr
11	Sankofa	Kofi Ghanaba
12	Sankofa	Oyafunmike Ogunlano
13	Sankofa	Alexandra Duah
14	Sankofa	Nick Medley

```
dfm2.info()#has total of 44946 actors

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 44947 entries, 0 to 44946
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  ---
0    title   44947 non-null    object
1    actor   44947 non-null    object
dtypes: object(2)
memory usage: 702.4+ KB

col3 = df_movie['listed_in'].apply(lambda x:str(x).split(',')).tolist()
dfm3 = pd.DataFrame(col3,index=df_movie['title'])
dfm3 = dfm3.stack()#transpose
dfm3 = pd.DataFrame(dfm3.reset_index())
dfm3.rename(columns={0: 'listed_in'},inplace=True)
dfm3.drop(columns='level_1',inplace=True)
dfm3.head()
#gives title of movie and their genres
```

	title	listed_in
0	Dick Johnson Is Dead	Documentaries
1	My Little Pony: A New Generation	Children & Family Movies
2	Sankofa	Dramas
3	Sankofa	Independent Movies
4	Sankofa	International Movies

```
#these are the different genres of the movies in the dataframe
dfm3['listed_in'].value_counts()
```

International Movies	2624
Dramas	1600
Comedies	1210
Action & Adventure	859
Documentaries	829
Dramas	827
Independent Movies	736
Romantic Movies	613
Children & Family Movies	605
Thrillers	512
Comedies	464
Music & Musicals	357
Stand-Up Comedy	334
Horror Movies	275
Sci-Fi & Fantasy	230
Sports Movies	218

```

International Movies      128
LGBTQ Movies              101
Horror Movies             82
Classic Movies            80
Faith & Spirituality      65
Thrillers                 65
Cult Movies               59
Movies                   54
Anime Features            50
Documentaries             40
Children & Family Movies  36
Classic Movies            36
Anime Features            21
Independent Movies        20
Music & Musicals           18
Sci-Fi & Fantasy           13
Cult Movies               12
Stand-Up Comedy           9
Romantic Movies           3
LGBTQ Movies              1
Sports Movies             1
Name: listed_in, dtype: int64

```

```

col4 = df_movie['country'].apply(lambda x:str(x).split(',')).tolist()
dfm4 = pd.DataFrame(col4,index=df_movie['title'])
dfm4 = dfm4.stack()#transpose
dfm4 = pd.DataFrame(dfm4.reset_index())
dfm4.rename(columns={0:'country'},inplace=True)
dfm4.drop(columns='level_1',inplace=True)
dfm4.head()
#gives title of movie and the country where it was produced

```

	title	country
0	Dick Johnson Is Dead	United States
1	My Little Pony: A New Generation	nan
2	Sankofa	United States
3	Sankofa	Ghana
4	Sankofa	Burkina Faso

```
dfm4['country'].value_counts()
```

```

United States    2361
India            927
nan              440
United States    388
United Kingdom   382
...
Bermuda          1
Ecuador          1
Iran              1
Armenia          1
Montenegro       1
Name: country, Length: 188, dtype: int64

```

```
df_movie['country'].value_counts()#if you look at this, US count is less than previous as we had not split the countries individually
```

```

United States    2055
India            893
United Kingdom   206
Canada           122
Spain            97
...
United Kingdom, Russia, United States    1
Paraguay, Argentina                     1
United Kingdom, Malawi                   1
Austria, Iraq, United States              1
United Arab Emirates, Jordan              1
Name: country, Length: 651, dtype: int64

```

```

#merging all the above dataframes based on common column ie title
dfm5 = dfm2.merge(dfm1,on=['title'],how='inner')#director&actor
dfm6 = dfm5.merge(dfm3,on=['title'],how='inner')#above_with_listed_in
dfm7 = dfm6.merge(dfm4,on=['title'],how='inner')#above_with_country
dfm7['actor'].replace(['nan'],['Unknown actor'],inplace=True)
dfm7['director'].replace(['nan'],['Unknown director'],inplace=True)
dfm7['country'].replace(['nan'],['NA'],inplace=True)
dfm7.head()

```

		title	actor	director	listed_in	country	
0		Dick Johnson Is Dead	Unknown actor	Kirsten Johnson	Documentaries	United States	
1		My Little Pony: A New Generation	Vanessa Hudgens	Robert Cullen	Children & Family Movies	NA	
2		My Little Pony: A New Generation	Vanessa Hudgens	José Luis Ucha	Children & Family Movies	NA	
3		My Little Pony: A New Generation	Kimiko Glenn	Robert Cullen	Children & Family Movies	NA	

```
df_movie.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90.0	Documentaries	As her father nears the end of his life, filmm...
2	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	NA	September 24, 2021	2021	PG	91.0	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...
3	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125.0	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s...

```
#merging this new df with original df_movie dataframe
df_movie_final = dfm7.merge(df_movie[['show_id', 'title', 'date_added', 'release_year', 'rating', 'duration', 'description']],on=['title'],how='left')
df_movie_final.head(5)
```

	title	actor	director	listed_in	country	show_id	date_added	release_year	rating	duration	description
0	Dick Johnson Is Dead	Unknown actor	Kirsten Johnson	Documentaries	United States	s1	September 25, 2021	2020	PG-13	90.0	As her father nears the end of his life, filmm...
1	My Little Pony: A New Generation	Vanessa Hudgens	Robert Cullen	Children & Family Movies	NA	s7	September 24, 2021	2021	PG	91.0	Equestria's divided. But a bright-eyed hero be...
2	My Little Pony: A New Generation	Vanessa Hudgens	José Luis Ucha	Children & Family Movies	NA	s7	September 24, 2021	2021	PG	91.0	Equestria's divided. But a bright-eyed hero be...
3	My Little Pony: A New Generation	Kimiko Glenn	Robert Cullen	Children & Family Movies	NA	s7	September 24, 2021	2021	PG	91.0	Equestria's divided. But a bright-eyed hero be...

```
#checking for nulls now
df_movie_final.isnull().sum()
```

title	0
actor	0
director	0
listed_in	0
country	0
show_id	0
date_added	0
release_year	0
rating	9
duration	0
description	0
dtype:	int64

```
df_movie_final[df_movie_final['rating'].isna()]
```

```

    title      actor  director listed_in country show_id date_added release_year rating duration description
0  88271  13TH: A Conversation with Oprah Winfrey & Ava ...  Oprah Winfrey  Unknown director  Movies  NA  s5990  January 26, 2017  2017  NaN  37.0  Oprah Winfrey sits down with director Ava DuVe...
1  88272  13TH: A Conversation with Oprah Winfrey & Ava ...  Ava DuVernay  Unknown director  Movies  NA  s5990  January 26, 2017  2017  NaN  37.0  Oprah Winfrey sits down with director Ava DuVe...
2  119939  My Honor Was Loyalty  Leone Frisa  Alessandro Pepe  Dramas  Italy  s7538  March 1, 2017  2015  NaN  115.0  Amid the chaos and horror of World War II, a c...
3  119939  My Honor Was Loyalty  Leone Frisa  Alessandro Pepe  Dramas  Italy  s7538  March 1, 2017  2015  NaN  115.0  Amid the chaos and horror of World War II, a c...

df_movie_final['rating'].fillna('NR',inplace=True)

df_movie_final.isnull().sum()

title      0
actor      0
director   0
listed_in  0
country    0
show_id    0
date_added 0
release_year 0
rating     0
duration   0
description 0
dtype: int64

df_movie_final.head(25)
```

	title	actor	director	listed_in	country	show_id	date_added	release_year	rating	duration	description
0	Dick Johnson Is Dead	Unknown actor	Kirsten Johnson	Documentaries	United States	s1	September 25, 2021	2020	PG-13	90.0	As her father nears the end of his life, filmm...
1	My Little Pony: A New Generation	Vanessa Hudgens	Robert Cullen	Children & Family Movies	NA	s7	September 24, 2021	2021	PG	91.0	Equestria's divided. But a bright-eyed hero be...
2	My Little Pony: A New Generation	Vanessa Hudgens	José Luis Ucha	Children & Family Movies	NA	s7	September 24, 2021	2021	PG	91.0	Equestria's divided. But a bright-eyed hero be...
3	My Little Pony: A New Generation	Kimiko Glenn	Robert Cullen	Children & Family Movies	NA	s7	September 24, 2021	2021	PG	91.0	Equestria's divided. But a bright-eyed hero be...
4	My Little Pony: A New Generation	Kimiko Glenn	José Luis Ucha	Children & Family Movies	NA	s7	September 24, 2021	2021	PG	91.0	Equestria's divided. But a bright-eyed hero be...
5	My Little Pony: A New Generation	James Marsden	Robert Cullen	Children & Family Movies	NA	s7	September 24, 2021	2021	PG	91.0	Equestria's divided. But a bright-eyed hero

```
df_movie_final['duration'].value_counts()
#this gives incorrect result as we have manipulated df,see below for actual duration of movies
```

```
94.0    4343
106.0    4040
97.0     3624
95.0     3560
96.0     3511
...
20.0         4
5.0          3
9.0          2
8.0          2
11.0         2
Name: duration, Length: 205, dtype: int64
```

Equestria's

```
df_movie['duration'].value_counts()
#using this to find number of movies ordered by their screentime in desc order
```

```
90.0     152
94.0     146
93.0     146
97.0     146
91.0     144
...
212.0      1
8.0         1
186.0      1
193.0      1
191.0      1
Name: duration, Length: 205, dtype: int64
```

Generation 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 20

```
<ipython-input-390-262eecdcc2c6>:1: FutureWarning: The default value of regex will change from True to False in a future version.
df_movie_final['title'] = df_movie_final['title'].str.replace(r"\(.*\)", "")

    title      actor  director  listed_in  country  show_id  date_added  release_year  rating  duration  description
0  Dick Johnson  Unknown  Kirsten  Documentaries  United  s1      September  2020  PG-13  90.0  As her father
   Is Dead      actor  Johnson                               States                               25, 2021                               nears the end of
                                                his life, filmm...

    title      actor  director  listed_in  country  show_id  date_added  release_year  rating  duration  description
1  My Little    Vanessa  Robert  Children &  NA      s7      September  2021  PG  91.0  Equestria's
   Pony: A New  ...      ...      Family Movies                               24, 2021                               divided. But a
                                                ...
```

Univariate Analysis on counts of each column

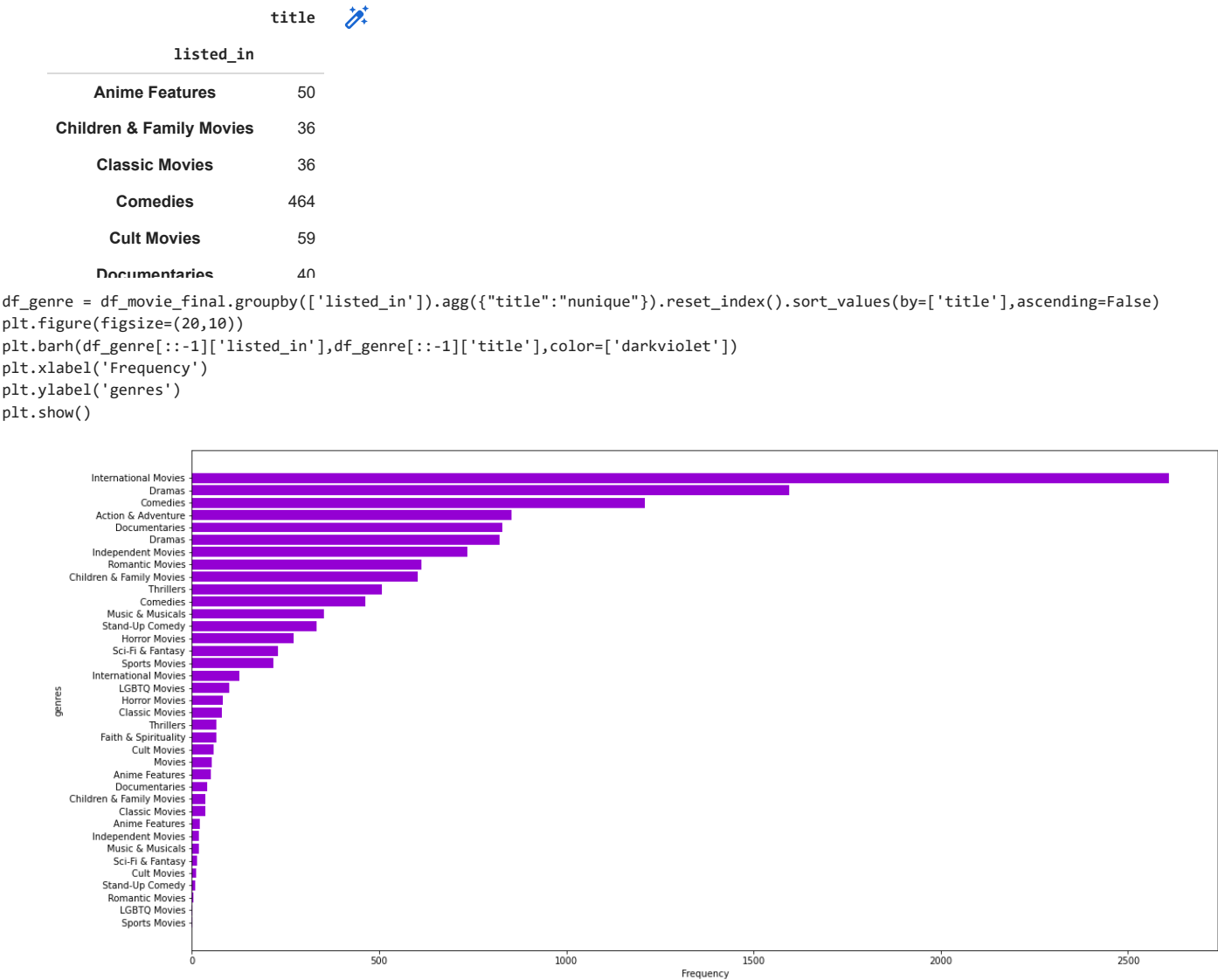
```
df_movie_final['listed_in'].value_counts

...
Equestria's

df_movie_final['listed_in'].value_counts

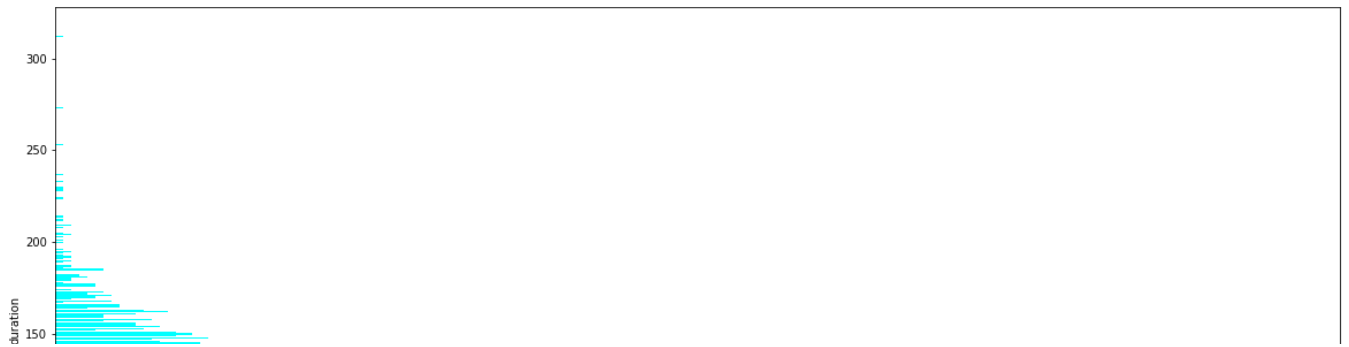
<bound method IndexOpsMixin.value_counts of 0
1      Children & Family Movies  Documentaries
2      Children & Family Movies
3      Children & Family Movies
4      Children & Family Movies
...
145909      International Movies
145910      Music & Musicals
145911      Dramas
145912      International Movies
145913      Music & Musicals
Name: listed_in, Length: 145914, dtype: object>

df_movie_final.groupby(['listed_in']).agg({"title": "nunique"})
```



LGBTQ Movies

1



Most of the movies are in the range 80-100 min and hence it can be said that the average length of the movie is somewhere around this range

```
df_international_movies_by_countries = df_movie_final.loc[df_movie_final['listed_in']=='International Movies']
df_international_movies_by_countries = df_international_movies_by_countries[['listed_in','country']]
df_international_movies_by_countries.head(40)
```

```

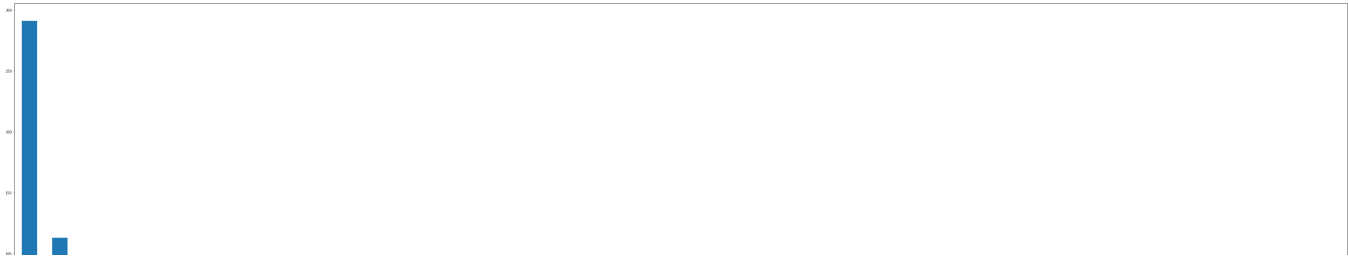
    listed_in  country
1278  International Movies    NA
1280  International Movies    NA
1282  International Movies    NA
1284  International Movies    NA

df_international_movies_by_countries.drop(df_international_movies_by_countries.loc[df_international_movies_by_countries['country']=='NA'])
df_international_movies_by_countries.reset_index(drop=True,inplace=True)
df_international_movies_by_countries.value_counts()
```

listed_in	country	
International Movies	India	291
	Spain	113
	Philippines	83
	United States	48
	South Korea	44
	Germany	35
	China	35
	Bulgaria	25
	United Arab Emirates	24
	Australia	21
	United Kingdom	18
	Germany	17
	Belgium	16
	Thailand	14
	Canada	13
	Italy	12
	France	12
	Nigeria	12
	Japan	11
	Poland	11
	Egypt	11
	Austria	10
	Romania	10
	Pakistan	10
	Kuwait	10
	Mexico	10
	Spain	9
	Canada	9
	South Africa	8
	France	8
	United Kingdom	8
	Denmark	8
	Netherlands	8
	Kenya	7
	Sweden	7
	Hong Kong	7
	Chile	6
	Taiwan	5
	Argentina	5
	Singapore	5
	Israel	4
	Romania	3
	Taiwan	3
	Netherlands	2

dtype: int64


```
df_international_movies_by_countries.country.value_counts().plot(kind='bar',figsize=(70,20))
xlabel='country'
ylabel= 'count of movies'
plt.xticks(fontsize=35)
plt.show()
#this shows that international movies are more popular in India followed by Spain and then Phillipines
#might be due to the reason that most international movies are in English and a huge population of India can understand English
```



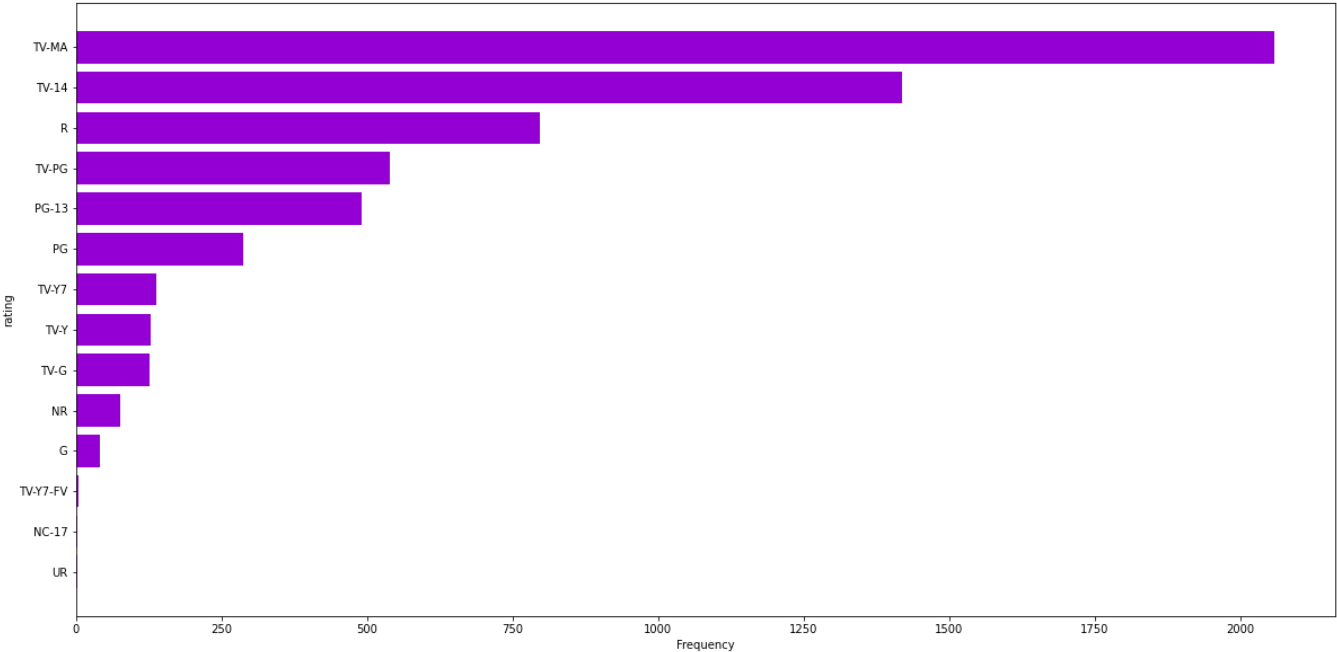
Similarly other sort of analysis can also be done



```
#number of distinct titles based on the basis of the country
df_22= df_movie_final.groupby(['country']).agg({'title':'nunique'})
df_22.tail(35)
```

title 	
country	
Netherlands	23
New Zealand	15
Nigeria	87
Norway	12
Pakistan	18
Paraguay	1
Peru	6
Philippines	77
Poland	22
Portugal	4
Romania	10
Russia	5
Saudi Arabia	7
Serbia	3
Singapore	10
Slovenia	1
Somalia	1
South Africa	35
South Korea	47
Soviet Union	1
Spain	129
Sweden	11
Switzerland	7
Syria	1
Taiwan	15
Thailand	41
Turkey	81
United Arab Emirates	20
United Kingdom	382
United States	2361
Uruguay	8
Venezuela	2
Vietnam	7
West Germany	1
Zimbabwe	1

```
df_rating = df_movie_final.groupby(['rating']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)
plt.figure(figsize=(20,10))
plt.barh(df_rating[0:-1]['rating'],df_rating[0:-1]['title'],color='darkviolet')
plt.xlabel('Frequency')
plt.ylabel('rating')
plt.show()
#TV-MA movies are most popular followed by TV-14
#these are the most viewed rating category
```



```
df_rating.value_counts()
```

```
rating    title
G          41      1
NC-17       3      1
NR          77      1
PG         287      1
PG-13      490      1
R          797      1
TV-14     1418      1
TV-G       126      1
TV-MA     2059      1
TV-PG      540      1
TV-Y       129      1
TV-Y7      139      1
TV-Y7-FV    5      1
UR          3      1
dtype: int64

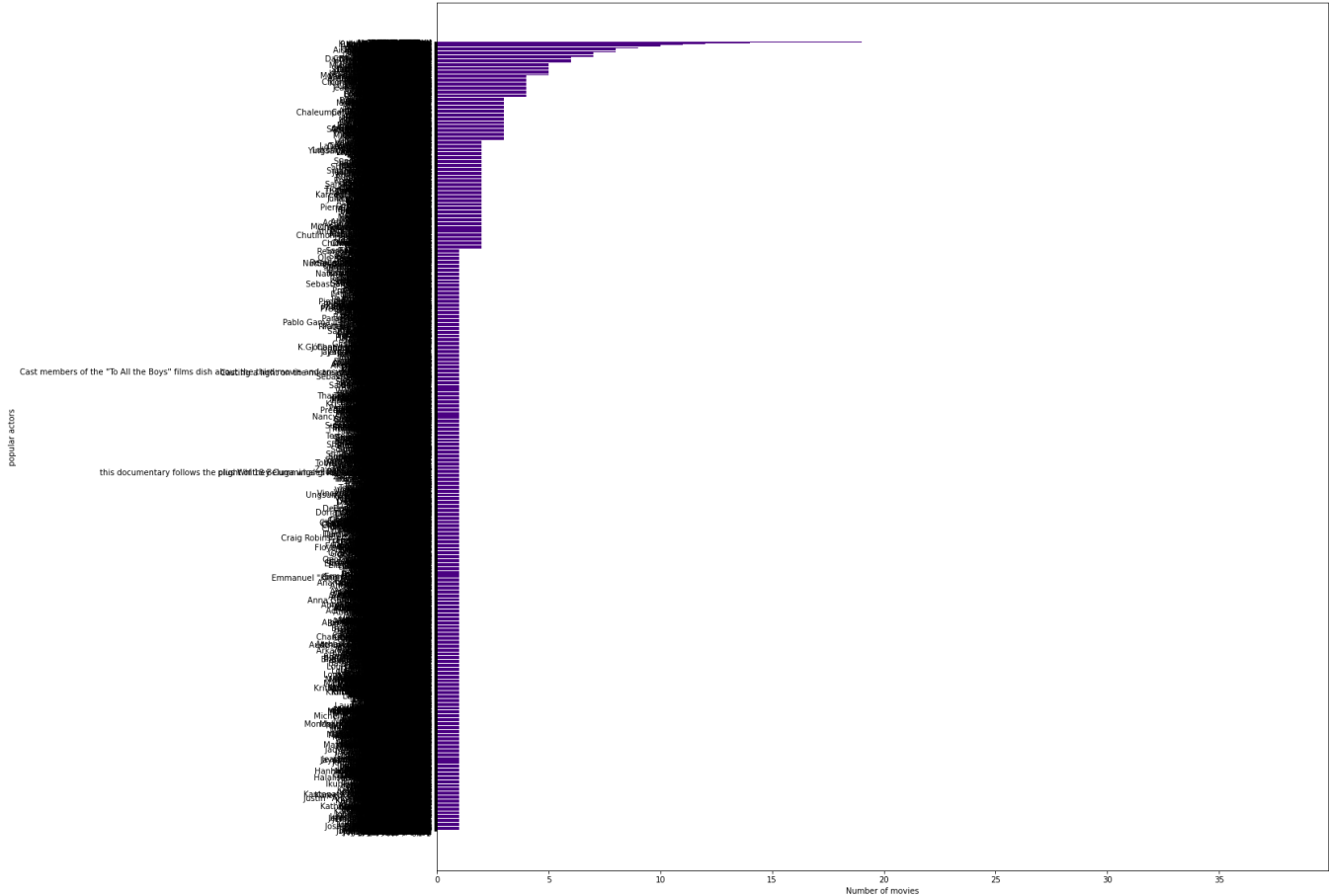
#number of distinct titles based on the actors
df_movie_actor = df_movie_final.groupby(['actor']).agg({'title':'nunique'})
df_movie_actor
```

title 	
actor	
Jr.	1

```
#renaming row from df having actor name as Joey Bada$$ as it was throwing an error
df_movie_final['actor'].replace({'Joey Bada$$':'Joey Badass'},inplace=True)

df_actor_count = df_movie_final.groupby(['actor']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)
df_actor_count = df_actor_count[df_actor_count['actor']!='Unknown actor']
plt.figure(figsize=(20,20))
plt.barh(df_actor_count[0:-1]['actor'],df_actor_count[0:-1]['title'],color=['indigo'])
plt.xlabel('Number of movies')
plt.ylabel('popular actors')
plt.show()

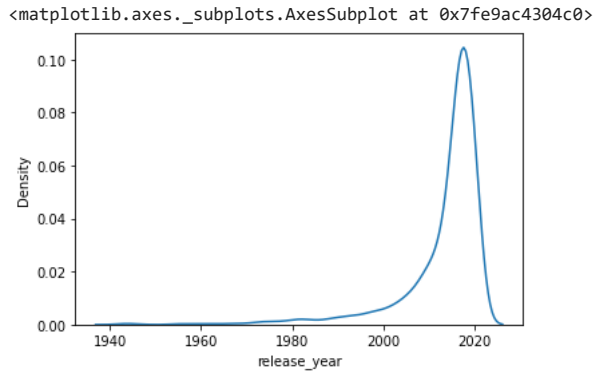
/usr/local/lib/python3.8/dist-packages/matplotlib/backends/backend_agg.py:214: RuntimeWarning: Glyph 12539 missing from current font
font.set_text(s, 0.0, flags=flags)
/usr/local/lib/python3.8/dist-packages/matplotlib/backends/backend_agg.py:183: RuntimeWarning: Glyph 12539 missing from current font
font.set_text(s, 0, flags=flags)
```



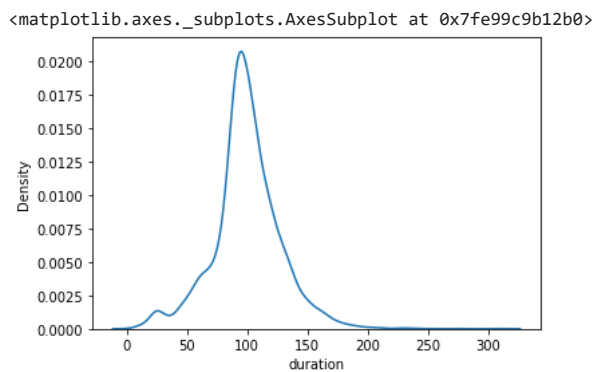
```
df_movie.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90.0	Documentaries	As her father nears the end of his life, filmm...
2	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	91.0	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...

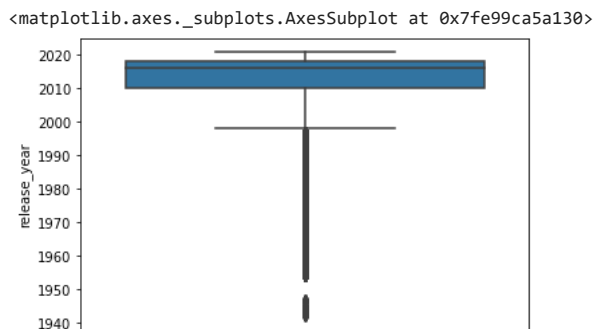
```
sns.kdeplot(df_movie['release_year'])
```



```
sns.kdeplot(df_movie['duration'])
```



```
sns.boxplot(y=df_movie_final['release_year'])
```



```
df_with_year = df_movie
df_with_year['year'] = pd.DatetimeIndex(df_with_year['date_added']).year
df_with_year.head()
```

```
<ipython-input-408-9cc435af5224>: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
df_with_year['year'] = pd.DatetimeIndex(df_with_year['date_added']).year
```

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90.0	Documentaries	As her father nears the end of his life, filmm...
2	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	91.0	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...

```
df_with_year_month = df_with_year
df_with_year_month['month'] = pd.DatetimeIndex(df_with_year_month['date_added']).month
df_with_year_month.head()
```

```
<ipython-input-409-28e87254e534>: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

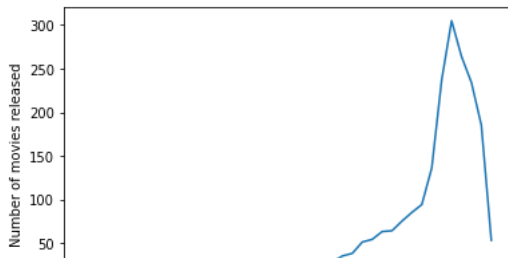
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus
df_with_year_month['month'] = pd.DatetimeIndex(df_with_year_month['date_added']).month
```

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90.0	Documentaries	As her father nears the end of his life, filmm...
2	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	91.0	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...
3	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D... Melissa McCarthy,	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125.0	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s...

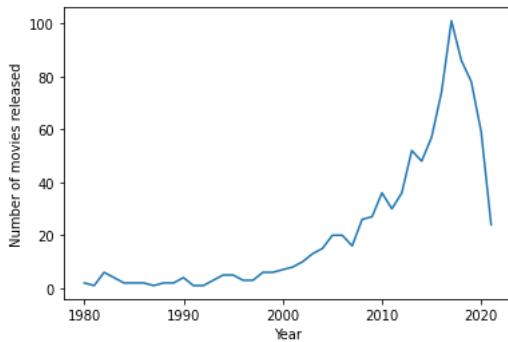
```
df_movie_US = df_movie_final.loc[df_movie_final['country']=='United States']
df_movie_US.head()
```

	title	actor	director	listed_in	country	show_id	date_added	release_year	rating	duration	description
0	Dick Johnson Is Dead	Unknown actor	Kirsten Johnson	Documentaries	United States	s1	September 25, 2021	2020	PG-13	90.0	As her father nears the end of his life, filmm...
21	Sankofa	Kofi Ghanaba	Haile Gerima	Dramas	United States	s8	September 24, 2021	1993	TV-MA	125.0	On a photo shoot in Ghana, an American model s...
27	Sankofa	Kofi Ghanaba	Haile Gerima	Independent Movies	United States	s8	September 24, 2021	1993	TV-MA	125.0	On a photo shoot in Ghana, an American model s...
33	Sankofa	Kofi Ghanaba	Haile Gerima	International Movies	United States	s8	September 24, 2021	1993	TV-MA	125.0	On a photo shoot in Ghana, an American model s...

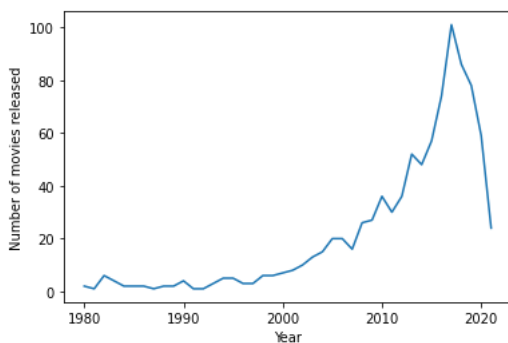
```
df_movie_US_release_year = df_movie_US[df_movie_US['release_year']>=1980].groupby(['release_year']).agg({'title':'nunique'}).reset_index()
sns.lineplot(data=df_movie_US_release_year,x='release_year',y='title')
plt.ylabel("Number of movies released")
plt.xlabel("Year")
plt.show()
```



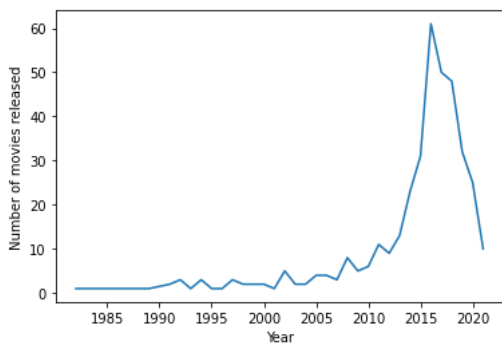
```
df_movie_India = df_movie_final.loc[df_movie_final['country']=='India']
df_movie_India_release_year = df_movie_India[df_movie_India['release_year']>=1980].groupby(['release_year']).agg({'title':'nunique'}).reset_index()
sns.lineplot(data=df_movie_India_release_year,x='release_year',y='title')
plt.ylabel("Number of movies released")
plt.xlabel("Year")
plt.show()
```



```
df_movie_India = df_movie_final.loc[df_movie_final['country']=='India']
df_movie_India_release_year = df_movie_India[df_movie_India['release_year']>=1980].groupby(['release_year']).agg({'title':'nunique'}).reset_index()
sns.lineplot(data=df_movie_India_release_year,x='release_year',y='title')
plt.ylabel("Number of movies released")
plt.xlabel("Year")
plt.show()
```



```
df_movie_UK = df_movie_final.loc[df_movie_final['country']=='United Kingdom']
df_movie_UK_release_year = df_movie_UK[df_movie_UK['release_year']>=1980].groupby(['release_year']).agg({'title':'nunique'}).reset_index()
sns.lineplot(data=df_movie_UK_release_year,x='release_year',y='title')
plt.ylabel("Number of movies released")
plt.xlabel("Year")
plt.show()
```



```
df_movie_US['release_year'].value_counts()
```

```
2017    3996
2018    3617
```



```
2016    3033
2019    2902
2020    2457
...
1945     12
1947     8
1942     6
1943     5
1946     2
Name: release_year, Length: 68, dtype: int64
```

```
#Let's do the same analysis for TV shows also
df_show.head()
```

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
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	Jailbirds New	NaN	NaN	NaN	September	2021	TV-MA	1 Season	Docuseries,	Feuds, flirtations and

```
df_show.nunique()

show_id      2676
type          1
title        2676
director     222
cast         2283
country      196
date_added   1051
release_year  46
rating        9
duration     15
listed_in    236
description  2672
dtype: int64
```

```
df_show.isnull().sum()

show_id      0
type         0
title        0
director    2446
cast        350
country     391
date_added   10
release_year  0
rating       2
duration     0
listed_in    0
description  0
dtype: int64
```

```
df_show.dropna(subset=['rating'], how='all', inplace=True)

/usr/local/lib/python3.8/dist-packages/pandas/util/_decorators.py:311: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
return func(*args, **kwargs)
```

```
df_show.isnull().sum()

show_id      0
type         0
title        0
director    2444
cast        350
country     391
date_added   10
release_year  0
```

```
rating      0
duration    0
listed_in   0
description 0
dtype: int64

df_show.head()

  show_id  type  title  director  cast  country  date_added  release_year  rating  duration  listed_in  description
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  Jailbirds New  NaN  NaN  NaN  September 2021  TV-MA  1 Season  Docuseries,  Feuds, flirtations and

df_show['country'].value_counts()

United States      760
United Kingdom     213
Japan              168
South Korea        158
India              79
...
Belarus            1
United Kingdom, Australia 1
France, Australia, Germany 1
Australia, New Zealand, United States 1
United States, France, South Korea, Indonesia 1
Name: country, Length: 196, dtype: int64

df_show['duration'] = df_show['duration'].str.replace("Seasons" , '')
df_show['duration'] = df_show['duration'].str.replace("Season" , '')
df_show.head()

<ipython-input-423-f3e793ce3087>:1: 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
df_show['duration'] = df_show['duration'].str.replace("Seasons" , '')
<ipython-input-423-f3e793ce3087>: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
df_show['duration'] = df_show['duration'].str.replace("Season" , '')

  show_id  type  title  director  cast  country  date_added  release_year  rating  duration  listed_in  description
1      s2  TV Show  Blood & Water  NaN  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...  South Africa  September 24, 2021  2021  TV-MA  2  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  Crime TV Shows, International TV Shows, TV Act...  To protect his family from a powerful drug lor...
3      s4  TV  Jailbirds New  NaN  NaN  NaN  September 2021  TV-MA  1  Docuseries,  Feuds, flirtations and

df_show['date_added'].fillna('NA',inplace=True)

/usr/local/lib/python3.8/dist-packages/pandas/core/generic.py:6392: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus
return self._update_inplace(result)
```

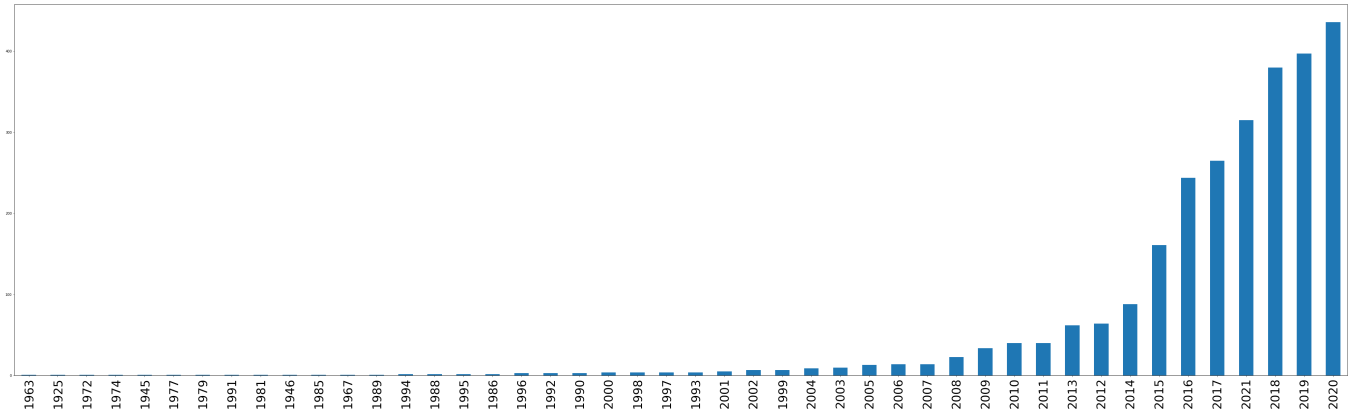


```
df_show.head(45)
```

show_id		type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descripti
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2	International TV Shows, TV Dramas, TV Mysteries	After crossi paths a party, a Ca Town
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1	Crime TV Shows, International TV Shows, TV Act...	To protect l family from powerful dr lo
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1	Docuseries, Reality TV	Feur flirtations a toilet talk down amc
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2	International TV Shows, Romantic TV Shows, TV ...	In a city coachi cente known to tr
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1	TV Dramas, TV Horror, TV Mysteries	The arrival a charisma young pri brir
6	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...	United Kingdom	September 24, 2021	2021	TV-14	9	British TV Shows, Reality TV	A talent batch amate bakers fa off ir
7	s11	TV Show	Vendetta: Truth, Lies and The Mafia	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1	Crime TV Shows, Docuseries, International TV S...	Sicity boa: a bold "Ar Maf coalition. E
8	s12	TV Show	Bangkok Breaking	Kongkiat Komesiri	Sukollawat Kanarot, Sushar Manaying, Pavarit M...	NaN	September 23, 2021	2021	TV-MA	1	Crime TV Shows, International TV Shows, TV Act...	Struggling earn a livi in Bangkok man
9	s15	TV Show	Crime Stories: India Detectives	NaN	NaN	NaN	September 22, 2021	2021	TV-MA	1	British TV Shows, Crime TV Shows, Docuseries	Camer followi Bengal police on t job
10	s16	TV Show	Dear White People	NaN	Logan Browning, Brandon P. Bell, DeRon Horton,...	United States	September 22, 2021	2021	TV-MA	4	TV Comedies, TV Dramas	Students color navig: the de slights z
11	s18	TV Show	Falsa identidad	NaN	Luis Ernesto Franco, Camila Sodi, Sergio Goyri...	Mexico	September 22, 2021	2020	TV-MA	2	Crime TV Shows, Spanish-Language TV Shows, TV ...	Strange Diego a Isabel fl their home
12	s20	TV Show	Jaguar	NaN	Blanca Suárez, Iván Marcos, Óscar Casas, Adriá...	NaN	September 22, 2021	2021	TV-MA	1	International TV Shows, Spanish-Language TV Sh...	In the 196l a Holocat survivor joi a gre
13	s21	TV Show	Monsters Inside: The 24 Faces of Billy Milligan	Olivier Megaton	NaN	NaN	September 22, 2021	2021	TV-14	1	Crime TV Shows, Docuseries, International TV S...	In the l: 1970s, accus serial rap c

```
df_show['release_year'].value_counts()
```

```
2020    436
2019    397
2018    380
2021    315
2017    265
2016    244
2015    161
2014     88
2012     64
2013     62
2010     40
2011     40
2009     34
2008     23
2006     14
2007     14
2005     13
2003     10
2004      9
1999      7
2002      7
2001      5
1993      4
2000      4
1997      4
1998      4
1990      3
1996      3
1992      3
1995      2
1994      2
1988      2
1986      2
1989      1
1967      1
1985      1
1946      1
1981      1
1972      1
1979      1
1977      1
1991      1
1974      1
1925      1
1945      1
1963      1
Name: release_year, dtype: int64
22      s41  Sci-Fi & Fantasy the Masters of the Universe  NaN  1981-1984  1981-1984  1981-1984  2021  TV-Y7  1  Sci-Fi & Fantasy
df_show[['release_year']].release_year.value_counts(ascending=True).plot(kind='bar', figsize=(70,20))
plt.xticks(fontsize=35)
plt.show()
#TV shows have also increased yearwise
```



Unnesting director,cast,country and listed_in

```
cola = df_show['director'].apply(lambda x:str(x).split(',')).tolist()
dfs1 = pd.DataFrame(colas,index=df_show['title'])
dfs1 = dfs1.stack()
dfs1 = pd.DataFrame(dfs1.reset_index())
dfs1.rename(columns={0:'director'},inplace=True)
```

```
dfs1.drop(columns='level_1',inplace=True)
dfs1.head()
```

	title	director
0	Blood & Water	nan
1	Ganglands	Julien Leclercq
2	Jailbirds New Orleans	nan
3	Kota Factory	nan
4	Midnight Mass	Mike Flanagan

```
colb = df_show['cast'].apply(lambda x:str(x).split(',')).tolist()
dfs2 = pd.DataFrame(colb,index=df_show['title'])
dfs2 = dfs2.stack()
dfs2 = pd.DataFrame(dfs2.reset_index())
dfs2.rename(columns={0: 'actor'},inplace=True)
dfs2.drop(columns='level_1',inplace=True)
dfs2.head()
```

	title	actor
0	Blood & Water	Ama Qamata
1	Blood & Water	Khosi Ngema
2	Blood & Water	Gail Mabalane
3	Blood & Water	Thabang Molaba
4	Blood & Water	Dillon Windvogel

```
colc = df_show['country'].apply(lambda x:str(x).split(',')).tolist()
dfs3 = pd.DataFrame(colc,index=df_show['title'])
dfs3 = dfs3.stack()
dfs3 = pd.DataFrame(dfs3.reset_index())
dfs3.rename(columns={0: 'country'},inplace=True)
dfs3.drop(columns='level_1',inplace=True)
dfs3.head()
```

	title	country
0	Blood & Water	South Africa
1	Ganglands	nan
2	Jailbirds New Orleans	nan
3	Kota Factory	India
4	Midnight Mass	nan

```
cold = df_show['listed_in'].apply(lambda x:str(x).split(',')).tolist()
dfs4 = pd.DataFrame(cold,index=df_show['title'])
dfs4 = dfs4.stack()
dfs4 = pd.DataFrame(dfs4.reset_index())
dfs4.rename(columns={0: 'listed_in'},inplace=True)
dfs4.drop(columns='level_1',inplace=True)
dfs4.head()
```

	title	listed_in
0	Blood & Water	International TV Shows
1	Blood & Water	TV Dramas
2	Blood & Water	TV Mysteries
3	Ganglands	Crime TV Shows
4	Ganglands	International TV Shows

```
#now we have to merge the above dataframes on the column name title
#merging all the above dataframes based on common column ie title
dfs5 = dfs2.merge(dfs1,on=['title'],how='inner')#director&actor
dfs6 = dfs5.merge(dfs3,on=['title'],how='inner')#above_with_listed_in
dfs7 = dfs6.merge(dfs4,on=['title'],how='inner')#above_with_country
dfs7['actor'].replace(['nan'],['Unknown actor'],inplace=True)
dfs7['director'].replace(['nan'],['Unknown director'],inplace=True)
dfs7['country'].replace(['nan'],['NA'],inplace=True)
dfs7.head()
```

	title	actor	director	country	listed_in
0	Blood & Water	Ama Qamata	Unknown director	South Africa	International TV Shows
1	Blood & Water	Ama Qamata	Unknown director	South Africa	TV Dramas
2	Blood & Water	Ama Qamata	Unknown director	South Africa	TV Mysteries
3	Blood & Water	Khosi Ngema	Unknown director	South Africa	International TV Shows
4	Blood & Water	Khosi Ngema	Unknown director	South Africa	TV Dramas



```
#merging this new df with original df_movie dataframe
df_show_final = dfs7.merge(df_show[['show_id','title','date_added','release_year','rating','duration','description']],on=['title'],how='left')
df_show_final.head(45)
```

	title	actor	director	country	listed_in	show_id	date_added	release_year	rating	duration	description
0	Blood & Water	Ama Qamata	Unknown director	South Africa	International TV Shows	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
1	Blood & Water	Ama Qamata	Unknown director	South Africa	TV Dramas	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
2	Blood & Water	Ama Qamata	Unknown director	South Africa	TV Mysteries	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
3	Blood & Water	Khosi Ngema	Unknown director	South Africa	International TV Shows	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
4	Blood & Water	Khosi Ngema	Unknown director	South Africa	TV Dramas	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
5	Blood & Water	Khosi Ngema	Unknown director	South Africa	TV Mysteries	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
6	Blood & Water	Gail Mabalane	Unknown director	South Africa	International TV Shows	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
7	Blood & Water	Gail Mabalane	Unknown director	South Africa	TV Dramas	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
8	Blood & Water	Gail Mabalane	Unknown director	South Africa	TV Mysteries	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
9	Blood & Water	Thabang Molaba	Unknown director	South Africa	International TV Shows	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
10	Blood & Water	Thabang Molaba	Unknown director	South Africa	TV Dramas	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
11	Blood & Water	Thabang Molaba	Unknown director	South Africa	TV Mysteries	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
12	Blood & Water	Dillon Windvogel	Unknown director	South Africa	International TV Shows	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
13	Blood & Water	Dillon Windvogel	Unknown director	South Africa	TV Dramas	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
14	Blood & Water	Dillon Windvogel	Unknown director	South Africa	TV Mysteries	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...

```
df_show_final.isnull().sum()
#phew, there are no nulls
```

title	0
actor	0
director	0
country	0
listed_in	0
show_id	0
date_added	0
release_year	0
rating	0
duration	0
description	0
dtype: int64	

Blood &

Unknown

South

September

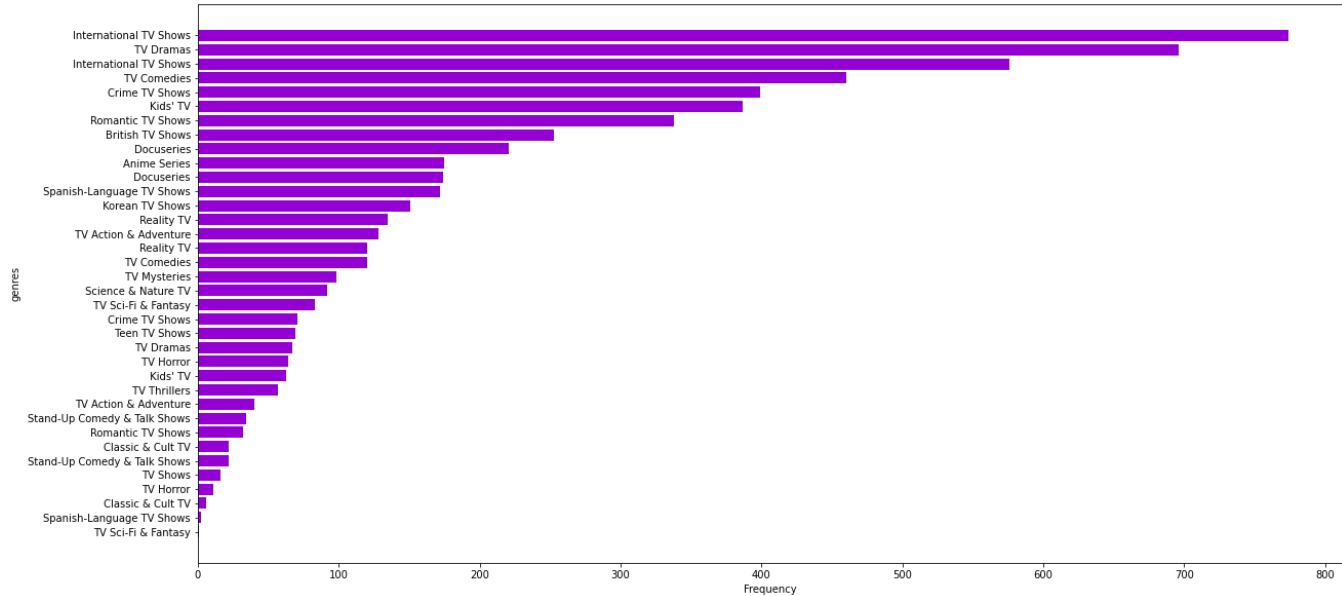
After crossing paths

Univariate Analysis on counts of each column

After crossing paths

Sum of international TV shows is maximum , hence investing or streaming this type of shows will increase revenue

```
df_genre = df_show_final.groupby(['listed_in']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)
plt.figure(figsize=(20,10))
plt.barh(df_genre[:-1]['listed_in'],df_genre[:-1]['title'],color=['darkviolet'])
plt.xlabel('Frequency')
plt.ylabel('genres')
plt.show()
```

```
df_international_shows_by_countries = df_show_final.loc[df_show_final['listed_in']=='International TV Shows']
df_international_shows_by_countries = df_international_shows_by_countries[['listed_in','country']]
df_international_shows_by_countries.head(40)
```

	listed_in	country
0	International TV Shows	South Africa
3	International TV Shows	South Africa
6	International TV Shows	South Africa
9	International TV Shows	South Africa
12	International TV Shows	South Africa
15	International TV Shows	South Africa
18	International TV Shows	South Africa
21	International TV Shows	South Africa
24	International TV Shows	South Africa
27	International TV Shows	South Africa
30	International TV Shows	South Africa
33	International TV Shows	South Africa
36	International TV Shows	South Africa

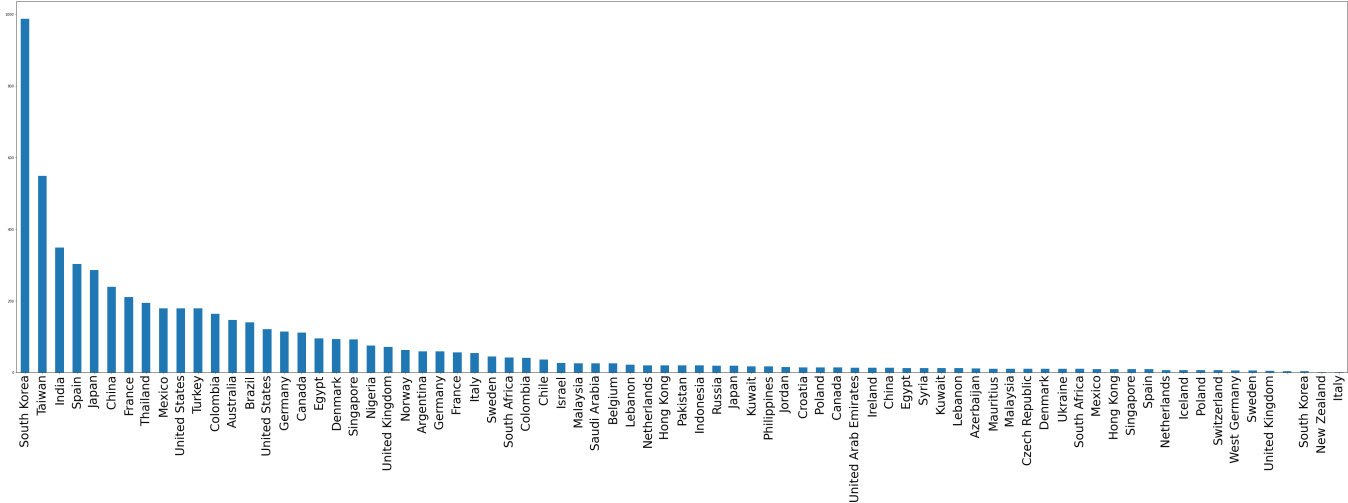
```
df_international_shows_by_countries.drop(df_international_shows_by_countries.loc[df_international_shows_by_countries['country']=='NA'],ir
df_international_shows_by_countries.reset_index(drop=True,inplace=True)
df_international_shows_by_countries.value_counts()
```

listed_in	country	
International TV Shows	South Korea	987
	Taiwan	549
	India	349
	Spain	303
	Japan	286
	...	
	United Kingdom	5
	South Korea	4
		4
	New Zealand	1
	Italy	1

Length: 77, dtype: int64

38 International TV Shows India

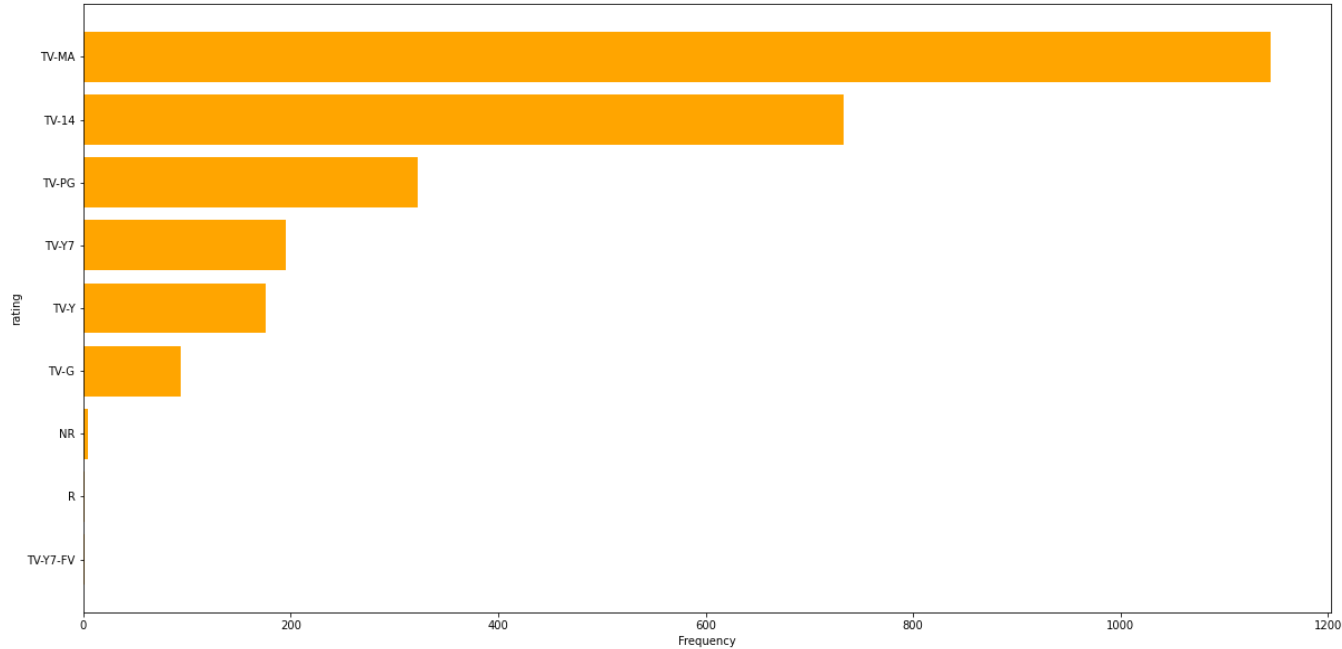
```
df_international_shows_by_countries.country.value_counts().plot(kind='bar',figsize=(70,20))
xlabel='country'
ylabel='count of shows'
plt.xticks(fontsize=35)
plt.show()
#this shows that international tv shows are more popular in South Korea followed by Taiwan and then India
```



```
#number of distinct titles based on the basis of the country
df_33= df_show_final.groupby(['country']).agg({'title':'nunique'})
df_33.tail(10)
```

	title
country	
Sweden	8
Switzerland	1
Taiwan	70
Thailand	24
Turkey	30
Ukraine	2
United Arab Emirates	1
United Kingdom	246
United States	847
Uruguay	1

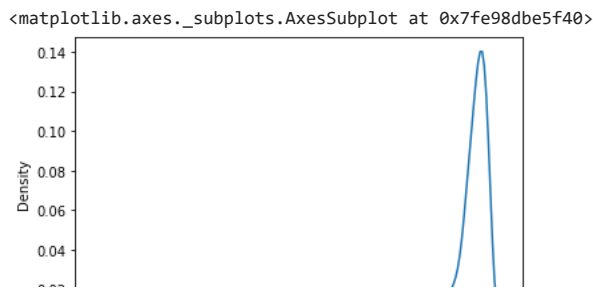
```
df_rating1 = df_show_final.groupby(['rating']).agg({"title":"nunique")).reset_index().sort_values(by=['title'],ascending=False)
plt.figure(figsize=(20,10))
plt.barh(df_rating1[::1]['rating'],df_rating1[::1]['title'],color=['orange'])
plt.xlabel('Frequency')
plt.ylabel('rating')
plt.show()
#TV-MA shows are most popular followed by TV-14
#these are the most viewed rating category
```



```
df_rating1.value_counts()

rating    title
NR         5      1
R          2      1
TV-14     733      1
TV-G       94      1
TV-MA    1145      1
TV-PG     323      1
TV-Y      176      1
TV-Y7     195      1
TV-Y7-FV   1       1
dtype: int64
```

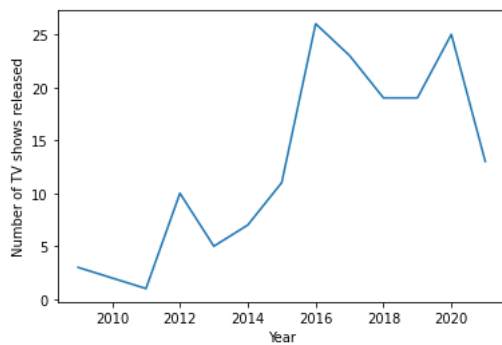
```
sns.kdeplot(df_show['release_year'])
```



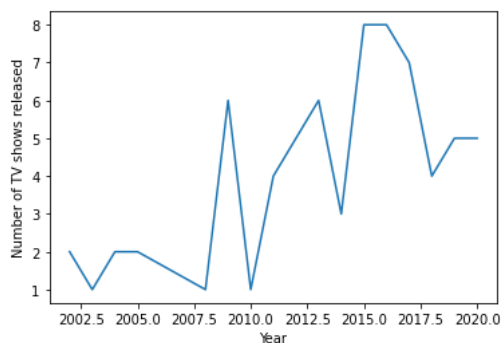
```
df_show_final['country'].value_counts()
```

```
United States    11552
NA               5698
Japan            4629
United Kingdom   4078
South Korea      3648
...
Russia           4
New Zealand       4
Uruguay           3
Israel            2
Czech Republic   1
Name: country, Length: 103, dtype: int64
```

```
df_show_SK = df_show_final.loc[df_show_final['country']=='South Korea']
df_movie_SK_release_year = df_show_SK[df_show_SK['release_year']>=1980].groupby(['release_year']).agg({'title':'nunique'}).reset_index()
sns.lineplot(data=df_movie_SK_release_year,x='release_year',y='title')
plt.ylabel("Number of TV shows released")
plt.xlabel("Year")
plt.show()
```



```
df_show_TW = df_show_final.loc[df_show_final['country']=='Taiwan']
df_movie_TW_release_year = df_show_TW[df_show_TW['release_year']>=1980].groupby(['release_year']).agg({'title':'nunique'}).reset_index()
sns.lineplot(data=df_movie_TW_release_year,x='release_year',y='title')
plt.ylabel("Number of TV shows released")
plt.xlabel("Year")
plt.show()
```



```
df_show_India = df_show_final.loc[df_show_final['country']=='India']
df_movie_India_release_year = df_show_India[df_show_India['release_year']>=1980].groupby(['release_year']).agg({'title':'nunique'}).reset_index()
sns.lineplot(data=df_movie_India_release_year,x='release_year',y='title')
plt.ylabel("Number of TV shows released")
plt.xlabel("Year")
plt.show()
```



df_show_final.head()

	title	actor	director	country	listed_in	show_id	date_added	release_year	rating	duration	description
0	Blood & Water	Ama Qamata	Unknown director	South Africa	International TV Shows	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
1	Blood & Water	Ama Qamata	Unknown director	South Africa	TV Dramas	s2	September 24, 2021	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...
2	Blood &	Ama	Unknown	South	TV Mysteries	s2	September	2021	TV-MA	2	After crossing paths at a party, a Cape Town

df_show_final['actor'].value_counts()

Unknown actor	818
David Attenborough	82
Takahiro Sakurai	54
Yuki Kaji	41
Ai Kayano	39
...	
Blue Chapman	1
Kirikou S'hai Muldrow	1
Thierry Mercier	1
Matthew Géczy	1
Jason Bolden	1

Name: actor, Length: 15489, dtype: int64

```
df_show_actor_count = df_show_final.groupby(['actor']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)
df_show_actor_count = df_show_actor_count[df_show_actor_count['actor']!='Unknown actor']
plt.figure(figsize=(20,100))
plt.barh(df_show_actor_count[:::-1]['actor'],df_show_actor_count[:::-1]['title'],color=['indigo'])
plt.xlabel('Number of TV shows')
plt.ylabel('popular actors')
plt.show()
```

[illegible]



[illegible]



✓

0s

completed at 12:34 AM

×