

## #Defining Problem Statement and Analysing basic metrics

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

!gdown

[https://d2beiqkhq929f0.cloudfront.net/public\\_assets/assets/000/000/940/original/netflix.csv](https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/000/940/original/netflix.csv)

Downloading...

From:

[https://d2beiqkhq929f0.cloudfront.net/public\\_assets/assets/000/000/940/original/netflix.csv](https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/000/940/original/netflix.csv)

To: /content/netflix.csv

0% 0.00/3.40M [00:00<?, ?B/s] 15% 524k/3.40M [00:00<00:00, 4.47MB/s] 100% 3.40M/3.40M [00:00<00:00, 16.6MB/s]

```
df=pd.read_csv('netflix.csv')
df.head()
```

	show_id	type	title	director	\
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	
1	s2	TV Show	Blood & Water	NaN	
2	s3	TV Show	Ganglands	Julien Leclercq	
3	s4	TV Show	Jailbirds New Orleans	NaN	
4	s5	TV Show	Kota Factory	NaN	

		cast	country	\
0		NaN	United States	
1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...		South Africa	
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...		NaN	
3		NaN	NaN	
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...		India	

	date_added	release_year	rating	duration	\
0	September 25, 2021	2020	PG-13	90 min	
1	September 24, 2021	2021	TV-MA	2 Seasons	
2	September 24, 2021	2021	TV-MA	1 Season	
3	September 24, 2021	2021	TV-MA	1 Season	
4	September 24, 2021	2021	TV-MA	2 Seasons	

	listed_in	\
0	Documentaries	
1	International TV Shows, TV Dramas, TV Mysteries	
2	Crime TV Shows, International TV Shows, TV Act...	
3	Docuseries, Reality TV	
4	International TV Shows, Romantic TV Shows, TV ...	

description

```
0 As her father nears the end of his life, filmm...
1 After crossing paths at a party, a Cape Town t...
2 To protect his family from a powerful drug lor...
3 Feuds, flirtations and toilet talk go down amo...
4 In a city of coaching centers known to train I...
```

```
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)
```

```
df.size
```

```
105684
```

```
df.dtypes
```

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

```
dtype: object
```

```
#Observations on the shape of data,
#data types of all the attributes,
#conversion of categorical attributes to 'category' (If required),
#missing value detection, statistical summary
```

```
df.duplicated().sum()
```

```
0
```

```
df['director'].replace(np.nan, 'No Data', inplace=True)
```

```
df['cast'].replace(np.nan, 'No Data', inplace=True)
```

```
categorical_column = ['type']
```

```
df['categorical_column'] = df[categorical_column].astype('category')
```

```
df.head()
```

	show_id	type	title	director	\
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	
1	s2	TV Show	Blood & Water	NaN	
2	s3	TV Show	Ganglands	Julien Leclercq	
3	s4	TV Show	Jailbirds New Orleans	NaN	
4	s5	TV Show	Kota Factory	NaN	

	cast	country	\
0	NaN	United States	
1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	
3	NaN	NaN	
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	

	date_added	release_year	rating	duration	\
0	September 25, 2021	2020	PG-13	90 min	
1	September 24, 2021	2021	TV-MA	2 Seasons	
2	September 24, 2021	2021	TV-MA	1 Season	
3	September 24, 2021	2021	TV-MA	1 Season	
4	September 24, 2021	2021	TV-MA	2 Seasons	

	listed_in	\
0	Documentaries	
1	International TV Shows, TV Dramas, TV Mysteries	
2	Crime TV Shows, International TV Shows, TV Act...	
3	Docuseries, Reality TV	
4	International TV Shows, Romantic TV Shows, TV ...	

```
description
```

```
categorical_column
```

```
0 As her father nears the end of his life, filmm...
```

```
Movie
```

1	After crossing paths at a party, a Cape Town t...	TV Show
2	To protect his family from a powerful drug lor...	TV Show
3	Feuds, flirtations and toilet talk go down amo...	TV Show
4	In a city of coaching centers known to train I...	TV Show

#5

```
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
categorical_column  0
dtype: int64
```

```
df['director'].replace(np.nan, 'No Data', inplace=True)
df['cast'].replace(np.nan, 'No Data', inplace=True)
```

#5)

```
df.isna()
```

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

	release_year	rating	duration	listed_in	description	\
0	False	False	False	False	False	
1	False	False	False	False	False	
2	False	False	False	False	False	

3	False	False	False	False	False
4	False	False	False	False	False
...	...	...	...	...	...
8802	False	False	False	False	False
8803	False	False	False	False	False
8804	False	False	False	False	False
8805	False	False	False	False	False
8806	False	False	False	False	False

	categorical_column
0	False
1	False
2	False
3	False
4	False
...	...
8802	False
8803	False
8804	False
8805	False
8806	False

[8807 rows x 13 columns]

#5)

df.isnull().sum()

show_id	0
type	0
title	0
director	0
cast	0
country	831
date_added	10
release_year	0
rating	4
duration	3
listed_in	0
description	0
categorical_column	0

dtype: int64

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

df.head()

	show_id	type	title	director	\
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	
1	s2	TV Show	Blood & Water	No Data	
2	s3	TV Show	Ganglands	Julien Leclercq	
3	s4	TV Show	Jailbirds New Orleans	No Data	



	release_year	rating	duration	listed_in	description	\
0	False	False	False	False	False	
1	False	False	False	False	False	
2	False	False	False	False	False	
3	False	False	False	False	False	
4	False	False	False	False	False	
...	...	...	...	...	...	
8802	False	False	False	False	False	
8803	False	False	False	False	False	
8804	False	False	False	False	False	
8805	False	False	False	False	False	
8806	False	False	False	False	False	

	categorical_column
0	False
1	False
2	False
3	False
4	False
...	...
8802	False
8803	False
8804	False
8805	False
8806	False

[8807 rows x 13 columns]

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

show_id	0
type	0
title	0
director	0
cast	0
country	0
date_added	10
release_year	0
rating	4
duration	3
listed_in	0
description	0
categorical_column	0

dtype: int64

```
movie_rating = df.loc[df['type'] == 'Movie', 'rating'].mode()[0]
```

```
tv_rating = df.loc[df['type'] == 'TV Show', 'rating'].mode()[0]
```

```
df['rating'] = df.apply(lambda x: movie_rating if x['type'] == 'Movie'
```

```

and pd.isna(x['rating'])
else tv_rating if x['type'] == 'TV Show' and
pd.isna(x['rating'])
else x['rating'], axis=1)

```

```
df.isna()
```

	show_id	type	title	director	cast	country	date_added	\
0	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	
8802	False	False	False	False	False	False	False	
8803	False	False	False	False	False	False	False	
8804	False	False	False	False	False	False	False	
8805	False	False	False	False	False	False	False	
8806	False	False	False	False	False	False	False	

	release_year	rating	duration	listed_in	description	\
0	False	False	False	False	False	
1	False	False	False	False	False	
2	False	False	False	False	False	
3	False	False	False	False	False	
4	False	False	False	False	False	
...	...	...	...	...	...	
8802	False	False	False	False	False	
8803	False	False	False	False	False	
8804	False	False	False	False	False	
8805	False	False	False	False	False	
8806	False	False	False	False	False	

	categorical_column
0	False
1	False
2	False
3	False
4	False
...	...
8802	False
8803	False
8804	False
8805	False
8806	False

```
[8807 rows x 13 columns]
```

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



```

show_id      0
type         0
title        0
director     0
cast         0
country      0
date_added   10
release_year  0
rating       0
duration     3
listed_in    0
description   0
categorical_column  0
dtype: int64

```

```

movie_duration_mode = df.loc[df['type'] == 'Movie', 'duration'].mode()
[0]

```

```

tv_duration_mode = df.loc[df['type'] == 'TV Show', 'duration'].mode()
[0]

```

```

df['duration'] = df.apply(lambda x: movie_duration_mode if x['type']
== 'Movie'

```

```

                        and pd.isna(x['duration'])
                        else tv_duration_mode if x['type'] == 'TV
Show'

```

```

                        and pd.isna(x['duration'])
                        else x['duration'], axis=1)

```

```

df.isnull().sum()

```

```

show_id      0
type         0
title        0
director     0
cast         0
country      0
date_added   10
release_year  0
rating       0
duration     0
listed_in    0
description   0
categorical_column  0
dtype: int64

```

```

df.dropna(inplace=True)

```

```

df.isnull().sum()

```

```

show_id      0
type         0

```

```

title          0
director       0
cast           0
country        0
date_added     0
release_year   0
rating         0
duration       0
listed_in      0
description    0
categorical_column  0
dtype: int64

```

### #3) Non-Graphical Analysis: Value counts and unique attributes

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

```

Movie      6131
TV Show    2666
Name: type, dtype: int64

```

```
df['director'].unique()
```

```

array(['Kirsten Johnson', 'No Data', 'Julien Leclercq', ...,
       'Majid Al Ansari', 'Peter Hewitt', 'Mozes Singh'],
      dtype=object)

```

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

```

No Data          2624
Rajiv Chilaka    19
Raúl Campos, Jan Suter  18
Suhas Kadav      16
Marcus Raboy     16
...
Raymie Muzquiz, Stu Livingston  1
Joe Menendez          1
Eric Bross            1
Will Eisenberg       1
Mozes Singh           1
Name: director, Length: 4529, dtype: int64

```

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

```

United States    3642
India            972
United Kingdom   418
Japan            244
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['rating'].unique()

array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
      'TV-G', 'G', 'NC-17', '74 min', '84 min', '66 min', 'NR',
      'TV-Y7-FV', 'UR'], dtype=object)

df['release_year'].unique()

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

df['release_year'].value_counts()

2018      1146
2017      1032
2019      1030
2020       953
2016       901
...
1959         1
1925         1
1961         1
1947         1
1966         1
Name: release_year, Length: 74, dtype: int64

df['cast'].value_counts()

No Data
825
David Attenborough
19
Vatsal Dubey, Julie Tejjwani, Rupa Bhimani, Jigna Bhardwaj, Rajesh
Kava, Mousam, Swapnil

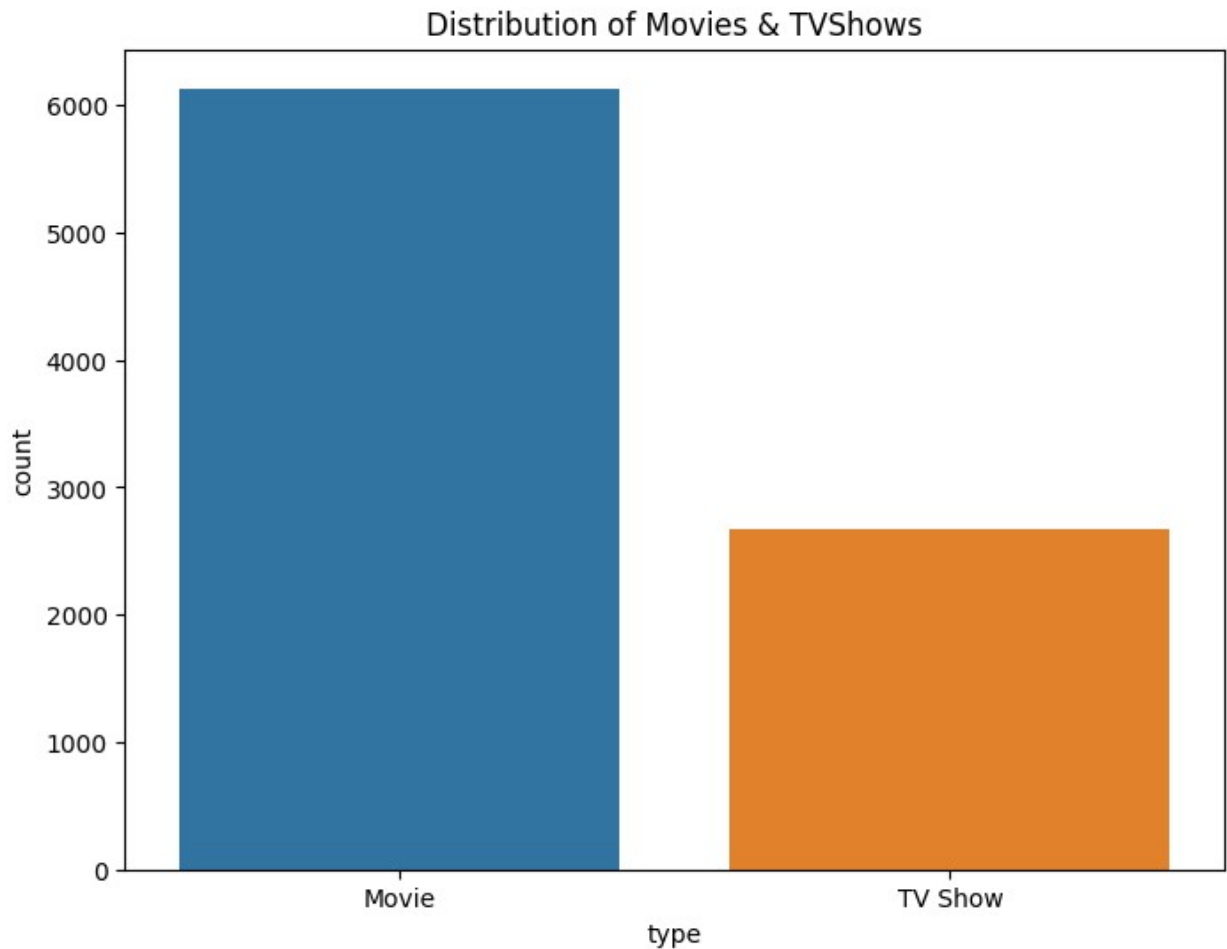
```

```
14
Samuel West
10
Jeff Dunham
7
```

```
...
Sanjay Dutt, Arjun Kapoor, Kriti Sanon, Zeenat Aman, Mohnish Bahl,
Padmini Kolhapure, Kunal Kapoor, Suhasini Mulay
1
Lika Berning, Bobby van Jaarsveld, Marlee van der Merwe, Sonja
Herholdt, Elize Cawood, Rouel Beukes, Kevin Leo, Paul du Toit,
Sylvaine Strike      1
Lisa Vicari, Dennis Mojen, Walid Al-Atiyat, Christina Hecke, Zoë
Straub, Lisa Hagmeister, Hans-Jochen Wagner, André Eisermann, Ernst
Stötzner      1
Piotr Cyrwus, Mikołaj Kubacki, Anna Radwan, Marian Dziędziel, Jan
Nowicki, Juliusz Chrzastowski, Małgorzata Krzysica, Jacek Strama
1
Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary,
Meghna Malik, Malkeet Rauni, Anita Shabdish, Chittaranjan Tripathy
1
Name: cast, Length: 7683, dtype: int64
```

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

```
plt.figure(figsize=(8,6))
sns.countplot(x='type',data=df)
plt.title('Distribution of Movies & TVShows')
plt.show()
```



```
df["date_added"] = pd.to_datetime(df['date_added'])
```

```
df.head()
```

	show_id	type	title	director \
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson
1	s2	TV Show	Blood & Water	No Data
2	s3	TV Show	Ganglands	Julien Leclercq
3	s4	TV Show	Jailbirds New Orleans	No Data
4	s5	TV Show	Kota Factory	No Data

	cast	country \
0	No Data	United States
1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	United States
3	No Data	United States
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India

	date_added	release_year	rating	duration \
0	2021-09-25	2020	PG-13	90 min
1	2021-09-24	2021	TV-MA	2 Seasons

```

2 2021-09-24      2021  TV-MA    1 Season
3 2021-09-24      2021  TV-MA    1 Season
4 2021-09-24      2021  TV-MA    2 Seasons

```

```

                                listed_in \
0                                Documentaries
1  International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act...
3                                Docuseries, Reality TV
4  International TV Shows, Romantic TV Shows, TV ...

```

```

                                description
categorical_column
0  As her father nears the end of his life, filmm...
Movie
1  After crossing paths at a party, a Cape Town t...      TV
Show
2  To protect his family from a powerful drug lor...      TV
Show
3  Feuds, flirtations and toilet talk go down amo...      TV
Show
4  In a city of coaching centers known to train I...      TV
Show

```

```
df.head()
```

```

  show_id  type      title      director \
0      s1  Movie  Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show      Blood & Water      No Data
2      s3  TV Show      Ganglands  Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans      No Data
4      s5  TV Show      Kota Factory      No Data

```

```

                                cast      country \
0                                No Data  United States
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...  South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...  United States
3                                No Data  United States
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...  India

```

```

  date_added  release_year  rating  duration \
0  2021-09-25      2020  PG-13      90 min
1  2021-09-24      2021  TV-MA    2 Seasons
2  2021-09-24      2021  TV-MA    1 Season
3  2021-09-24      2021  TV-MA    1 Season
4  2021-09-24      2021  TV-MA    2 Seasons

```

```

                                listed_in \
0                                Documentaries
1  International TV Shows, TV Dramas, TV Mysteries

```

```

2 Crime TV Shows, International TV Shows, TV Act...
3 Docuseries, Reality TV
4 International TV Shows, Romantic TV Shows, TV ...

```

```

description

```

```

categorical_column \

```

```

0 As her father nears the end of his life, filmm...

```

```

Movie

```

```

1 After crossing paths at a party, a Cape Town t...      TV
Show

```

```

2 To protect his family from a powerful drug lor...      TV
Show

```

```

3 Feuds, flirtations and toilet talk go down amo...      TV
Show

```

```

4 In a city of coaching centers known to train I...      TV
Show

```

```

month_added month_name_added year_added
0           9      September      2021
1           9      September      2021
2           9      September      2021
3           9      September      2021
4           9      September      2021

```

```

df['cast'] = df['cast'].str.split(', ')

```

```

df['director'] = df['director'].str.split(', ')

```

```

df['country'] = df['country'].str.split(', ')

```

```

df = df.explode('cast').explode('director').explode('country')

```

```

df.head()

```

```

show_id      type      title      director
cast \
0      s1      Movie      Dick Johnson Is Dead      Kirsten Johnson      No
Data
1      s2      TV Show      Blood & Water      No Data      Ama
Qamata
1      s2      TV Show      Blood & Water      No Data      Khosi
Ngema
1      s2      TV Show      Blood & Water      No Data      Gail
Mabalane
1      s2      TV Show      Blood & Water      No Data      Thabang
Molaba

```

```

country      date_added      release_year      rating      duration \
0      United States      2021-09-25      2020      PG-13      90 min
1      South Africa      2021-09-24      2021      TV-MA      2 Seasons
1      South Africa      2021-09-24      2021      TV-MA      2 Seasons
1      South Africa      2021-09-24      2021      TV-MA      2 Seasons

```

```

1   South Africa   2021-09-24           2021   TV-MA   2 Seasons

                                listed_in \
0                                Documentaries
1   International TV Shows, TV Dramas, TV Mysteries
1   International TV Shows, TV Dramas, TV Mysteries
1   International TV Shows, TV Dramas, TV Mysteries
1   International TV Shows, TV Dramas, TV Mysteries

                                description
categorical_column \
0   As her father nears the end of his life, filmm...
Movie
1   After crossing paths at a party, a Cape Town t...   TV
Show
1   After crossing paths at a party, a Cape Town t...   TV
Show
1   After crossing paths at a party, a Cape Town t...   TV
Show
1   After crossing paths at a party, a Cape Town t...   TV
Show

   month_added month_name_added year_added
0             9      September      2021
1             9      September      2021
1             9      September      2021
1             9      September      2021
1             9      September      2021

```

```

plt.figure(figsize=(8,6))
sns.distplot(df['release_year'],kde=False, bins=30, color='blue')
plt.title('Distribution of movies across release years')
plt.xlabel('release_year')
plt.ylabel('count')
plt.show()

```

<ipython-input-77-66f317104f83>:2: UserWarning:

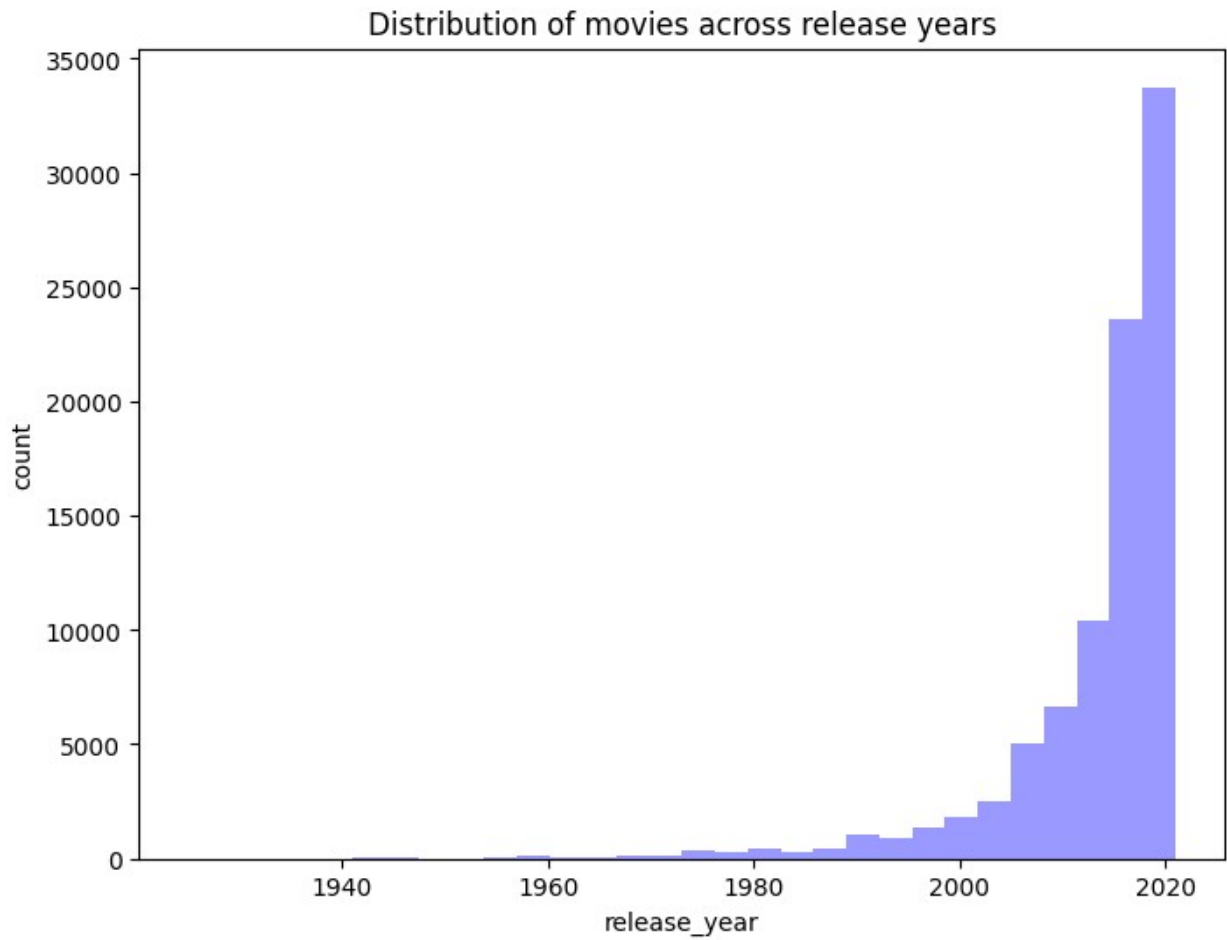
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

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

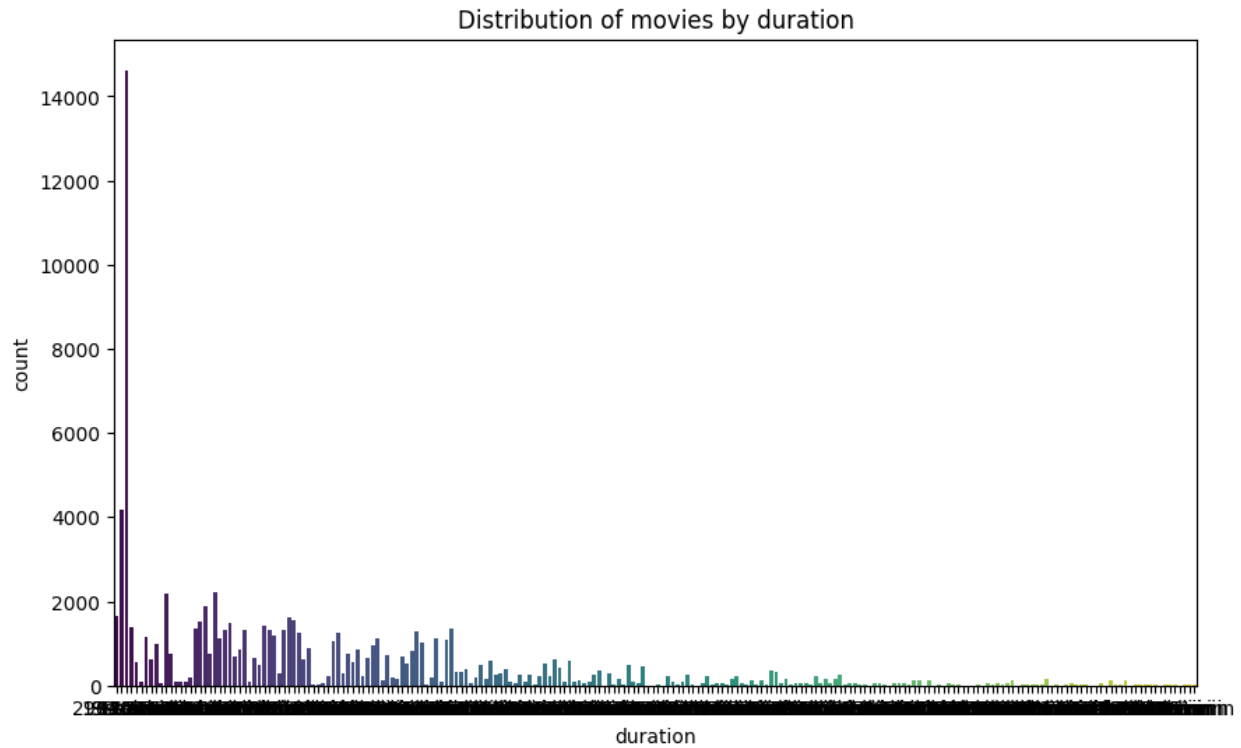
For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(df['release_year'],kde=False, bins=30, color='blue')
```

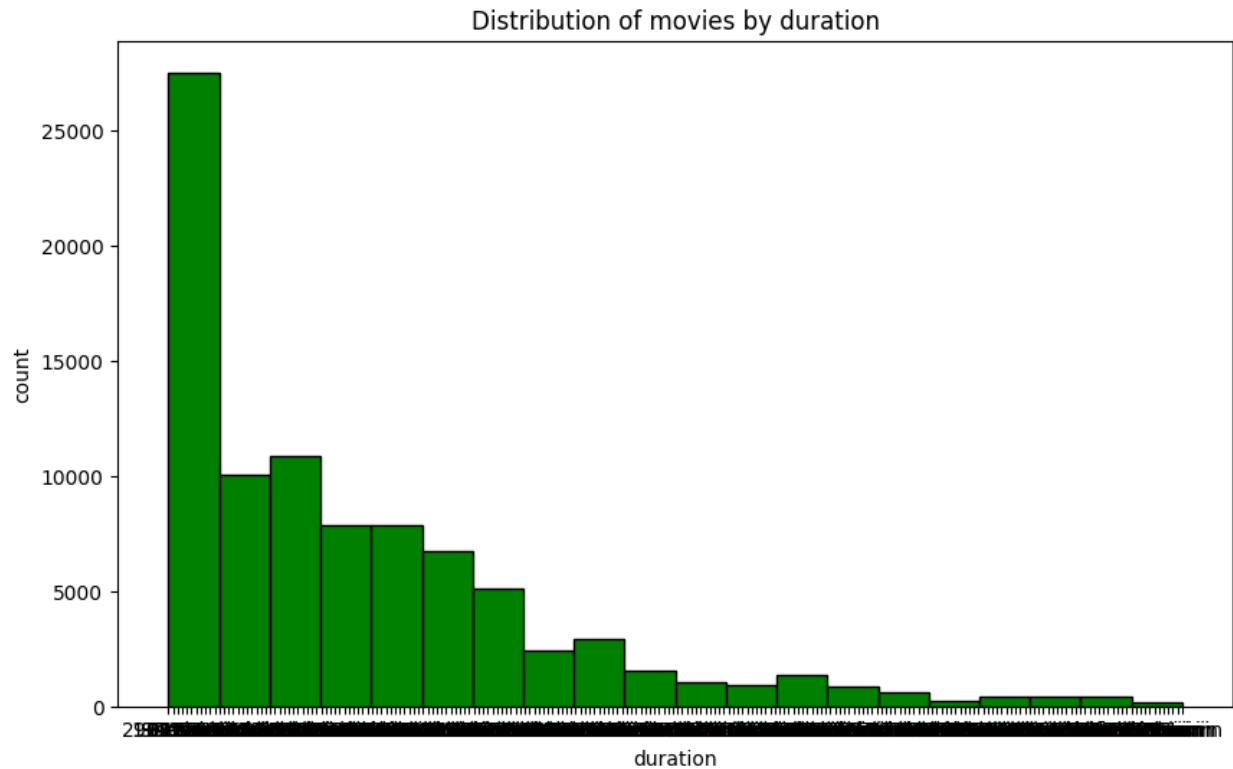


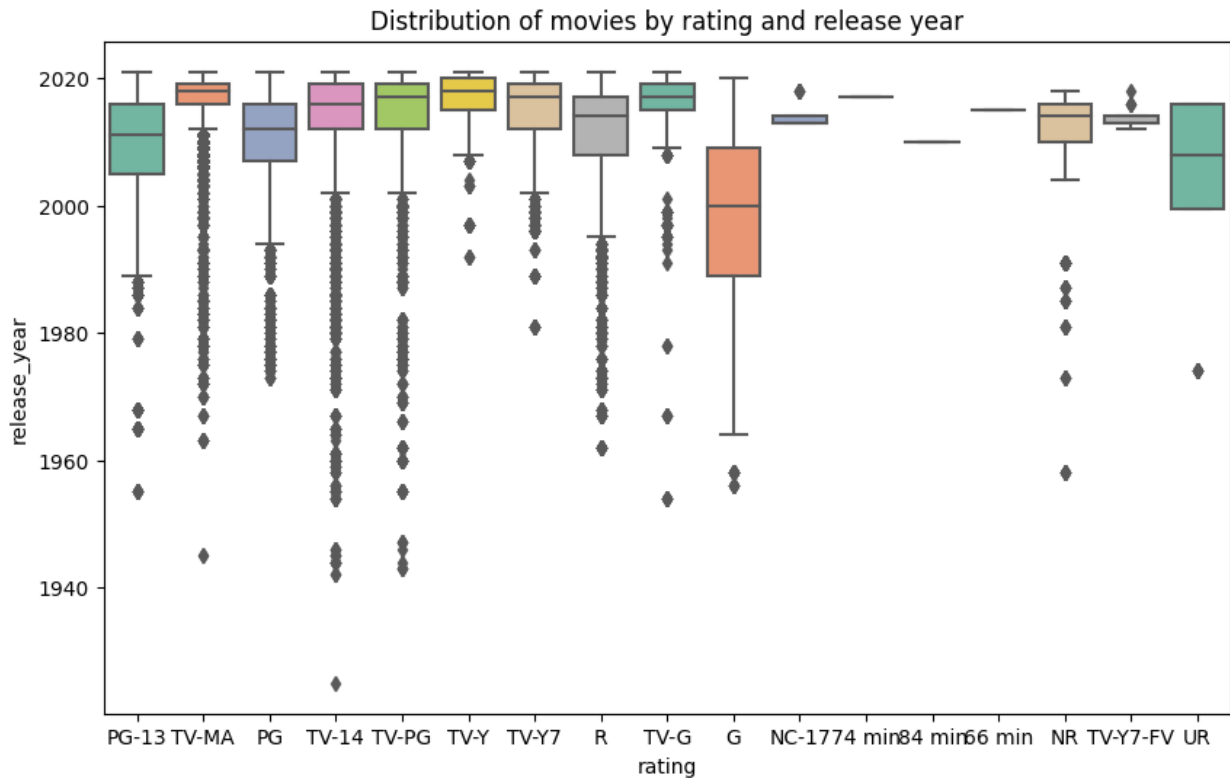


```
plt.figure(figsize=(10,6))
sns.countplot(x='duration', data=df, palette='viridis')
plt.title('Distribution of movies by duration')
plt.xlabel('duration')
plt.ylabel('count')
plt.show()
```



```
plt.figure(figsize=(10,6))
plt.hist(df['duration'], bins=20, color='green', edgecolor='black')
plt.title('Distribution of movies by duration')
plt.xlabel('duration')
plt.ylabel('count')
plt.show()
```





```
plt.figure(figsize=(10,6))
sns.heatmap(df[['release_year','cast']].corr(), annot=True,
cmap='coolwarm', fmt=".2f")
plt.title('Distribution of movies by rating and release year')
#plt.xlabel('rating')
#plt.ylabel('release_year')
plt.show()
```

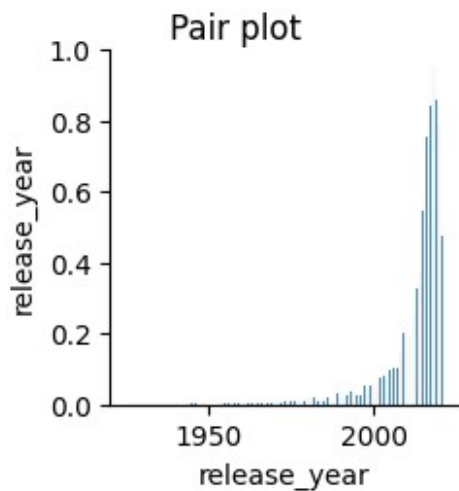
<ipython-input-96-d88930fefdf2>:2: FutureWarning: The default value of numeric\_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric\_only to silence this warning.

```
sns.heatmap(df[['release_year','cast']].corr(), annot=True,
cmap='coolwarm', fmt=".2f")
```



```
plt.figure(figsize=(10,6))
x=['release_year','duration']
sns.pairplot(df[x])
plt.suptitle('Pair plot',y=1.02)
plt.show()
```

<Figure size 1000x600 with 0 Axes>



```
df.head()
```

	show_id	type	title	director	\
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	
1	s2	TV Show	Blood & Water	NaN	
2	s3	TV Show	Ganglands	Julien Leclercq	
3	s4	TV Show	Jailbirds New Orleans	NaN	
4	s5	TV Show	Kota Factory	NaN	

	cast	country	\
0	NaN	United States	
1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	
3	NaN	NaN	
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	

	date_added	release_year	rating	duration	\
0	September 25, 2021	2020	PG-13	90 min	
1	September 24, 2021	2021	TV-MA	2 Seasons	
2	September 24, 2021	2021	TV-MA	1 Season	
3	September 24, 2021	2021	TV-MA	1 Season	
4	September 24, 2021	2021	TV-MA	2 Seasons	

	listed_in	\
0	Documentaries	
1	International TV Shows, TV Dramas, TV Mysteries	
2	Crime TV Shows, International TV Shows, TV Act...	
3	Docuseries, Reality TV	
4	International TV Shows, Romantic TV Shows, TV ...	

	description
0	As her father nears the end of his life, filmm...
1	After crossing paths at a party, a Cape Town t...
2	To protect his family from a powerful drug lor...
3	Feuds, flirtations and toilet talk go down amo...
4	In a city of coaching centers known to train I...

```
top3_ry = df['release_year'].value_counts().index[:3]
top3_rat = df['rating'].value_counts().index[:3]
#top3_tit = df['title'].value_counts().index[:3]

top3_df = df.loc[(df["rating"].isin(top3_rat)) &
(df["release_year"].isin(top3_ry))]
top3_df.head()
```

	show_id	type	title	\
21	s22	TV Show	Resurrection:	
73	s74	Movie	King of	

```

Boys
90      s91      Movie                                Paradise
Hills
96      s97      Movie  If I Leave Here Tomorrow: A Film About
Lynyrd ...
105     s106     Movie                                Angamaly
Diaries

                                director
cast \
21      NaN      Engin Altan Düzyatan, Serdar Gökhan, Hülya
Dar...
73      Kemi Adetiba  Sola Sobowale, Adesua Etomi, Remilekun
"Remini...
90      Alice Waddington  Emma Roberts, Danielle Macdonald,
Awkwafina, E...
96      Stephen Kijak  Ronnie Van Zandt, Gary Rossington, Allen
Colli...
105     Lijo Jose Pellissery  Antony Varghese, Reshma Rajan, Binny Rinky
Ben...

                                country          date_added  release_year  rating
duration \
21      Turkey      September 22, 2021          2018  TV-14  5
Seasons
73      Nigeria     September 14, 2021          2018  TV-MA
182 min
90      Spain, United States  September 9, 2021          2019  TV-MA
95 min
96      United States  September 7, 2021          2018  TV-MA
97 min
105     India        September 5, 2021          2017  TV-14
128 min

                                listed_in \
21      International TV Shows, TV Action & Adventure,...
73      Dramas, International Movies
90      Sci-Fi & Fantasy, Thrillers
96      Documentaries, Music & Musicals
105     Action & Adventure, Comedies, Dramas

                                description
21      When a good deed unwittingly endangers his cla...
73      When a powerful businesswoman's political ambi...
90      Uma wakes up in a lush tropical facility desig...
96      Using interviews and archival footage, this do...
105     After growing up amidst the gang wars of his h...

corr=top3_df.corr()
corr

```

```
<ipython-input-16-b672d6e930ae>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it
will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
```

```
corr=top3_df.corr()
```

```
           release_year
release_year           1.0
```

```
sns.heatmap(corr)
plt.show()
```



```
df.head()
```

	show_id	type	title	director	\
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	
1	s2	TV Show	Blood & Water	NaN	
2	s3	TV Show	Ganglands	Julien Leclercq	
3	s4	TV Show	Jailbirds New Orleans	NaN	
4	s5	TV Show	Kota Factory	NaN	

	cast	country	\
0	NaN	United States	
1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	



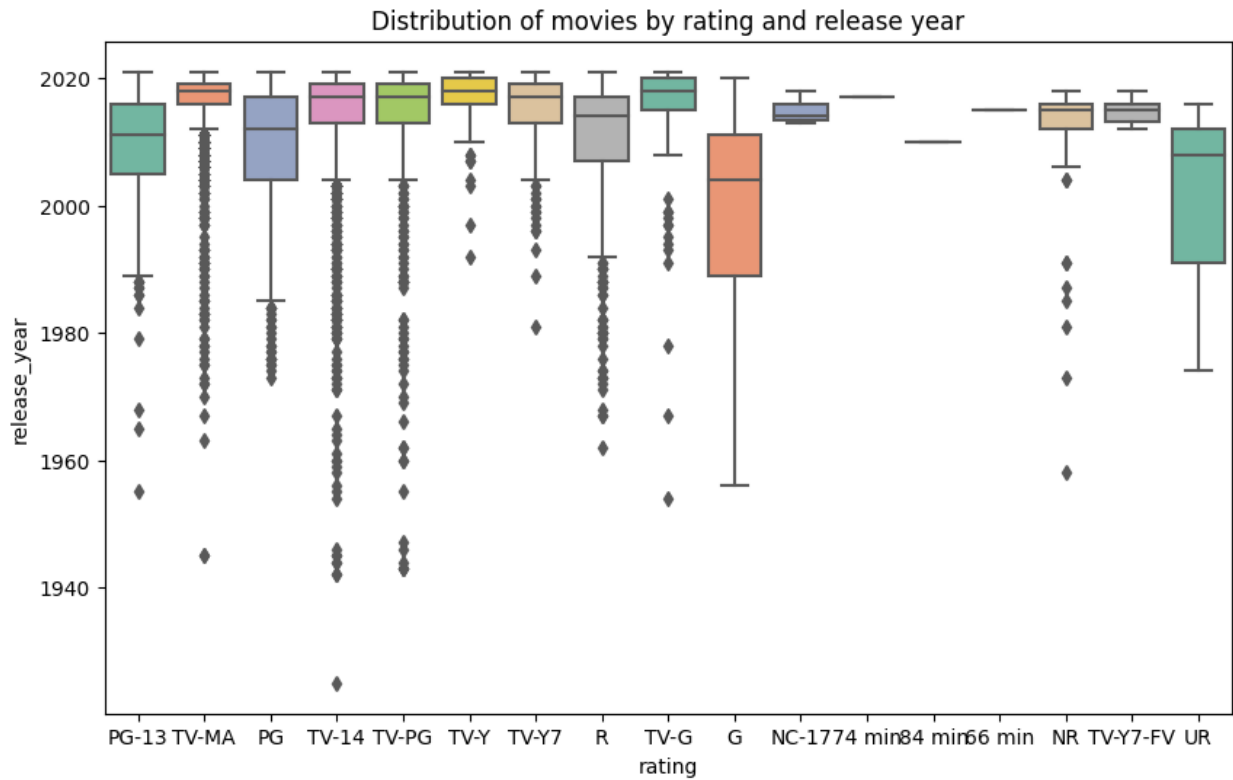
3			NaN		NaN
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...				India

	date_added	release_year	rating	duration	\
0	September 25, 2021	2020	PG-13	90 min	
1	September 24, 2021	2021	TV-MA	2 Seasons	
2	September 24, 2021	2021	TV-MA	1 Season	
3	September 24, 2021	2021	TV-MA	1 Season	
4	September 24, 2021	2021	TV-MA	2 Seasons	

	listed_in	\
0	Documentaries	
1	International TV Shows, TV Dramas, TV Mysteries	
2	Crime TV Shows, International TV Shows, TV Act...	
3	Docuseries, Reality TV	
4	International TV Shows, Romantic TV Shows, TV ...	

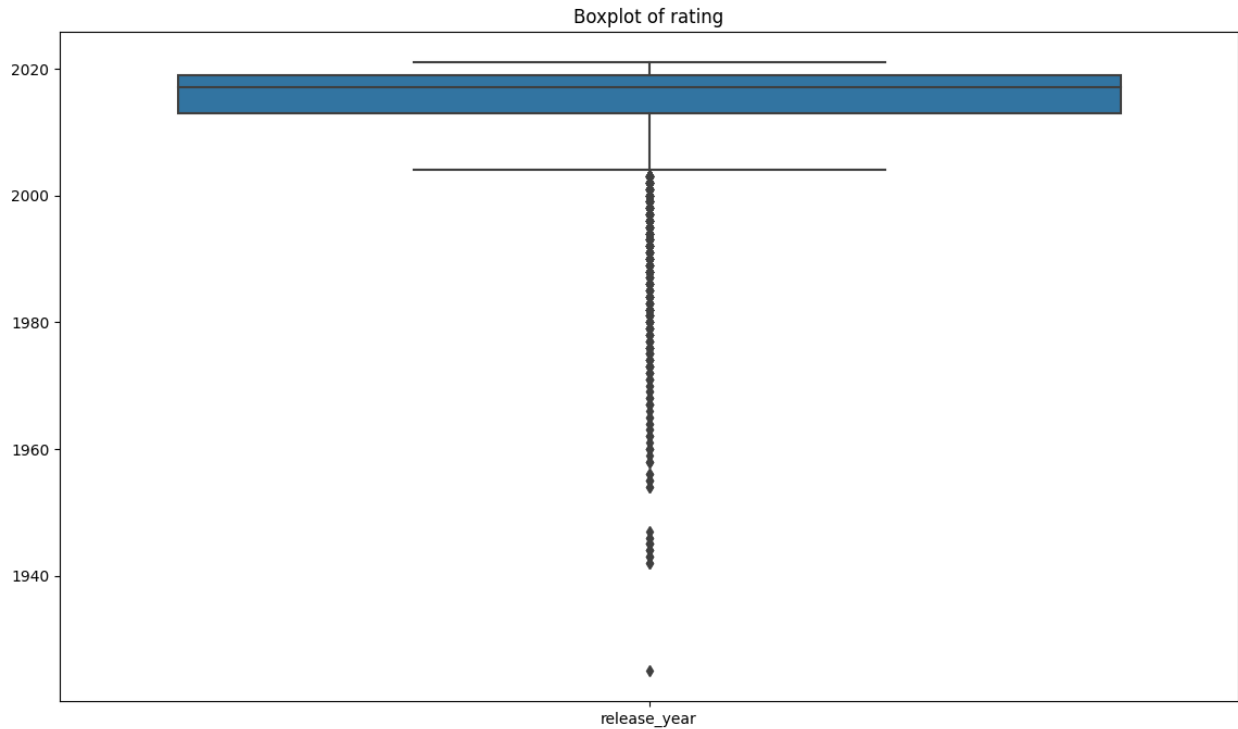
	description
0	As her father nears the end of his life, filmm...
1	After crossing paths at a party, a Cape Town t...
2	To protect his family from a powerful drug lor...
3	Feuds, flirtations and toilet talk go down amo...
4	In a city of coaching centers known to train I...

```
plt.figure(figsize=(10,6))
sns.boxplot(x='rating', y='release_year', data=df, palette='Set2')
plt.title('Distribution of movies by rating and release year')
plt.xlabel('rating')
plt.ylabel('release_year')
plt.show()
```

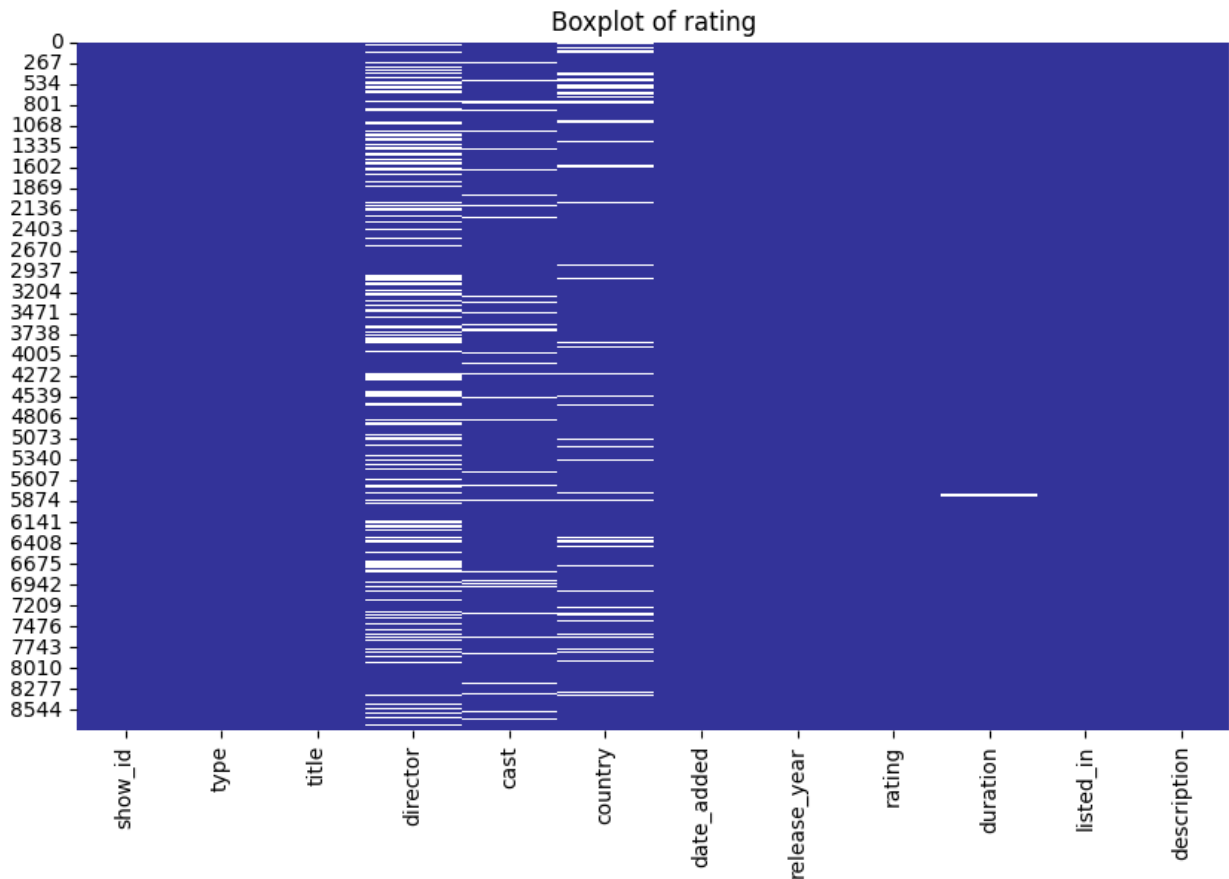


```
rating= df.select_dtypes(include=['number']).columns

plt.figure(figsize=(14, 8))
sns.boxplot(df[rating])
plt.title('Boxplot of rating')
plt.show()
```



```
#5) missing value in heatmap:  
plt.figure(figsize=(10, 6))  
sns.heatmap(df.isnull(),cbar=False,cmap='terrain')  
plt.title('Boxplot of rating')  
plt.show()
```



#6.1) Comments on the range of attributes

#Ans:

#Changed the date formation in given data

```
'''#we use attributes such as date changes
by using this function: df["date_added"] =
pd.to_datetime(df['date_added'])'''
```

'''#also we are used different types of plots  
such as box plot, pair plot, test plot, countplot, etc.,  
for analysing the data'''

#6.2) Comments on the distribution of the variables

#and relationship between them

#Ans:

'''#The given dataset netflix for distribution  
of the variables, I was used different plots  
to get understand of the relationship  
between different variables  
loaded in the dataset'''

#6.3) Comments for each univariate and bivariate plot

#Ans:

'''#The univariate plots different observations/values of the same variable corresponding to the index/observation number.'''

'''#Bivariate analysis is the simultaneous analysis of two variables. It explores the concept of the relationship between two variable whether there exists an association and the strength of this association or whether there are differences between two variables and the significance of these differenc'''

#7) Should include patterns observed in  
#the data along with what you can infer from it

'''Insights of the given Netflix dataset from the raw dataset, I was analysed and found many duplicates and some missing values in many columns and I removed the duplicates, replaced the missing values with \*NaN\* from the analysis, I get the below details such as top3 rating movies, top 3 directors and confirming that compared with TV series movies were got good rating  
'''

#8) Actionable items for business.  
#No technical jargon. No complications.  
#Simple action items that everyone can understand

'''Recommending to take or buy good rated movies to increase the business in different countries so here we can confirming that in some countries some kind of genres movies doing good, so we have to buy the related movies for the relaated audiences, then definitely we can increase our customer size'''