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

```python
df = pd.read_csv("/Users/mrunmay/desktop/netflix.csv")
```

```python
df
```

Out[3]:

	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, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...

8807 rows × 12 columns

```python
df.shape
```

Out[4]:

```
(8807, 12)
```

```python
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.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.describe()
```

	release_year
count	8807.000000
mean	2014.180198
std	8.819312
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

```
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
```

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

```
Movie     6131
TV Show   2676
Name: type, dtype: int64
```

```python
cast = df["cast"].apply(lambda x: str(x).split(", ")).tolist()
castdf = pd.DataFrame(cast,index=df["title"])
castnewdf=castdf.stack()
castnewdf=castnewdf.reset_index()
castnewdf
```

	title	level_1	0
0	Dick Johnson Is Dead	0	nan
1	Blood & Water	0	Ama Qamata
2	Blood & Water	1	Khosi Ngema
3	Blood & Water	2	Gail Mabalane
4	Blood & Water	3	Thabang Molaba
...	...	...	...
64946	Zubaan	3	Manish Chaudhary
64947	Zubaan	4	Meghna Malik
64948	Zubaan	5	Malkeet Rauni
64949	Zubaan	6	Anita Shabdish
64950	Zubaan	7	Chittaranjan Tripathy

64951 rows × 3 columns

```python
castnewdf.drop(columns="level_1",inplace = True)
```

```python
castnewdf
```

	title	0
0	Dick Johnson Is Dead	nan
1	Blood & Water	Ama Qamata
2	Blood & Water	Khosi Ngema
3	Blood & Water	Gail Mabalane
4	Blood & Water	Thabang Molaba
...	...	...
64946	Zubaan	Manish Chaudhary
64947	Zubaan	Meghna Malik
64948	Zubaan	Malkeet Rauni
64949	Zubaan	Anita Shabdish
64950	Zubaan	Chittaranjan Tripathy

64951 rows × 2 columns

```python
castnewdf.columns=["title","cast"]
```

```python
castnewdf.head()
```

	title	cast
0	Dick Johnson Is Dead	nan
1	Blood & Water	Ama Qamata
2	Blood & Water	Khosi Ngema
3	Blood & Water	Gail Mabalane
4	Blood & Water	Thabang Molaba

In [15]:

```
castnewdf.shape
```

Out[15]:

```
(64951, 2)
```

In [16]:

```
castnewdf.isna().sum()
```

Out[16]:

```
title    0
cast     0
dtype: int64
```

In [17]:

```
df_new = df.merge(castnewdf,on="title",how="left")
```

In [18]:

```
df_new
```

	show_id	type	title	director	cast_x	country	date_added	release_year	rating	duration	listed_in	description	cast_y
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...	nan
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...	Ama Qamata
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Khosi Ngema
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Gail Mabalane
4	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Thabang Molaba
...	...	...	...	...	...	...	...	...	...	...	...	...	...
64946	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	Manish Chaudhary
64947	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	Meghna Malik
64948	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	Malkeet Rauni
64949	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	Anita Shabdish
64950	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	Chittaranjan Tripathy

64951 rows × 13 columns

```python
df_new = df.drop(columns=["cast"]).merge(castnewdf,on="title",how="left")
```

```python
df_new
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in	description	cast
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	nan
1	s2	TV Show	Blood & Water	NaN	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...	Ama Qamata
2	s2	TV Show	Blood & Water	NaN	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...	Khosi Ngema
3	s2	TV Show	Blood & Water	NaN	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...	Gail Mabalane
4	s2	TV Show	Blood & Water	NaN	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...	Thabang Molaba
...	...	...	...	...	...	...	...	...	...	...	...	...
64946	s8807	Movie	Zubaan	Mozez Singh	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	Manish Chaudhary
64947	s8807	Movie	Zubaan	Mozez Singh	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	Meghna Malik
64948	s8807	Movie	Zubaan	Mozez Singh	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	Malkeet Rauni
64949	s8807	Movie	Zubaan	Mozez Singh	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	Anita Shabdish
64950	s8807	Movie	Zubaan	Mozez Singh	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	Chittaranjan Tripathy

64951 rows × 12 columns

```python
df_new.cast.value_counts().iloc[:5]
```

```
nan               825
Anupam Kher        43
Shah Rukh Khan     35
Julie Tejwani      33
Naseeruddin Shah   32
Name: cast, dtype: int64
```

```python
listeddate = df_new["listed_in"].apply(lambda x: str(x).split(", ")).tolist()
listeddate_new = pd.DataFrame(listeddate,index=df_new["title"])
listeddate_new=listeddate_new.stack()
listeddate_new=listeddate_new.reset_index()
listeddate_new
```

	title	level_1	0
0	Dick Johnson Is Dead	0	Documentaries
1	Blood & Water	0	International TV Shows
2	Blood & Water	1	TV Dramas
3	Blood & Water	2	TV Mysteries
4	Blood & Water	0	International TV Shows
...	...	...	...
149507	Zubaan	1	International Movies
149508	Zubaan	2	Music & Musicals
149509	Zubaan	0	Dramas
149510	Zubaan	1	International Movies
149511	Zubaan	2	Music & Musicals

149512 rows × 3 columns

listeddate_new**.**drop(columns**=**"level_1",inplace**=True**)

listeddate_new

	title	0
0	Dick Johnson Is Dead	Documentaries
1	Blood & Water	International TV Shows
2	Blood & Water	TV Dramas
3	Blood & Water	TV Mysteries
4	Blood & Water	International TV Shows
...	...	...
149507	Zubaan	International Movies
149508	Zubaan	Music & Musicals
149509	Zubaan	Dramas
149510	Zubaan	International Movies
149511	Zubaan	Music & Musicals

149512 rows × 2 columns

listeddate_new**.**columns**=**["title","listed_in"]

listeddate_new

	title	listed_in
0	Dick Johnson Is Dead	Documentaries
1	Blood & Water	International TV Shows
2	Blood & Water	TV Dramas
3	Blood & Water	TV Mysteries
4	Blood & Water	International TV Shows
...	...	...
149507	Zubaan	International Movies
149508	Zubaan	Music & Musicals
149509	Zubaan	Dramas
149510	Zubaan	International Movies
149511	Zubaan	Music & Musicals

149512 rows × 2 columns

```
df_final=df_new.drop(columns=["listed_in"]).merge(listeddate_new,on="title",how="left")
```

```
df_final.head()
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	description	cast	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	September 25, 2021	2020	PG-13	90 min	As her father nears the end of his life, filmm...	nan	Documentaries
1	s2	TV Show	Blood & Water	NaN	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Ama Qamata	International TV Shows
2	s2	TV Show	Blood & Water	NaN	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Ama Qamata	TV Dramas
3	s2	TV Show	Blood & Water	NaN	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Ama Qamata	TV Mysteries
4	s2	TV Show	Blood & Water	NaN	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Ama Qamata	International TV Shows

```
df_final.shape
```

(1559790, 12)

```
df_final.isnull().sum()
```

```
show_id            0
type               0
title              0
director      536723
country       110077
date_added      1604
release_year       0
rating          1119
duration           3
description        0
cast               0
listed_in          0
dtype: int64
```

```
df_final.dropna(inplace=True)
```

```
df_final.isnull().sum()
```

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

```python
df_final.shape
```

```
(962719, 12)
```

```python
df_final.drop_duplicates(inplace=True)
```

```python
df_final.nunique()
```

```
show_id      8807
type         2
title        8807
director     4528
country      748
date_added   1767
release_year 74
rating       17
duration     220
description  8775
cast         36440
listed_in    42
dtype: int64
```

```python
df_final.drop_duplicates()
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	description	cast	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	September 25, 2021	2020	PG-13	90 min	As her father nears the end of his life, filmm...	nan	Documentaries
2389	s8	Movie	Sankofa	Haile Gerima	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Kofi Ghanaba	Dramas
2390	s8	Movie	Sankofa	Haile Gerima	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Kofi Ghanaba	Independent Movies
2391	s8	Movie	Sankofa	Haile Gerima	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Kofi Ghanaba	International Movies
2413	s8	Movie	Sankofa	Haile Gerima	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Oyafunmike Ogunlano	Dramas
...	...	...	...	...	...	...	...	...	...	...	...	...
1559743	s8807	Movie	Zubaan	Mozez Singh	India	March 2, 2019	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Anita Shabdish	International Movies
1559744	s8807	Movie	Zubaan	Mozez Singh	India	March 2, 2019	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Anita Shabdish	Music & Musicals
1559766	s8807	Movie	Zubaan	Mozez Singh	India	March 2, 2019	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Chittaranjan Tripathy	Dramas
1559767	s8807	Movie	Zubaan	Mozez Singh	India	March 2, 2019	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Chittaranjan Tripathy	International Movies
1559768	s8807	Movie	Zubaan	Mozez Singh	India	March 2, 2019	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Chittaranjan Tripathy	Music & Musicals

98660 rows × 12 columns

```
df_final.shape
```

(98660, 12)

```
df_final["country"].value_counts()
```

```
United States                              28713
India                                      18844
United Kingdom                              2385
Japan                                       2309
Nigeria                                     2124
                                            ...
United States, Uruguay                         1
United States, Venezuela                       1
Germany, United States, Sweden                 1
United States, Denmark                         1
United States, Brazil, Japan, Spain, India     1
Name: country, Length: 658, dtype: int64
```

```
country = df_final["country"].apply(lambda x: str(x).split(", ")).tolist()
```

```
country = df_final["country"].apply(lambda x: str(x).split(",")).tolist()
country_new = pd.DataFrame(country,index=df_final["title"])
country_new=country_new.stack()
country_new=country_new.reset_index()
country_new
```

	title	level_1	0
0	Dick Johnson Is Dead	0	United States
1	Sankofa	0	United States
2	Sankofa	1	Ghana
3	Sankofa	2	Burkina Faso
4	Sankofa	3	United Kingdom
...	...	...	...
129442	Zubaan	0	India
129443	Zubaan	0	India
129444	Zubaan	0	India
129445	Zubaan	0	India
129446	Zubaan	0	India

129447 rows × 3 columns

```
country_new.drop(columns="level_1",inplace=True)
```

```
country_new.head()
```

	title	0
0	Dick Johnson Is Dead	United States
1	Sankofa	United States
2	Sankofa	Ghana
3	Sankofa	Burkina Faso
4	Sankofa	United Kingdom

```
country_new.columns=["title","country"]
```

```
country_new
```

	title	country
0	Dick Johnson Is Dead	United States
1	Sankofa	United States
2	Sankofa	Ghana
3	Sankofa	Burkina Faso
4	Sankofa	United Kingdom
...	...	...
129442	Zubaan	India
129443	Zubaan	India
129444	Zubaan	India
129445	Zubaan	India
129446	Zubaan	India

129447 rows × 2 columns

```
country_new.shape
```

```
(129447, 2)
```

```
final_data=df_final.drop(columns=["country"]).merge(country_new,on="title",how="left")
```

```
final_data.head()
```

	show_id	type	title	director	date_added	release_year	rating	duration	description	cast	listed_in	country
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	September 25, 2021	2020	PG-13	90 min	As her father nears the end of his life, filmm...	nan	Documentaries	United States
1	s8	Movie	Sankofa	Haile Gerima	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Kofi Ghanaba	Dramas	United States
2	s8	Movie	Sankofa	Haile Gerima	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Kofi Ghanaba	Dramas	Ghana
3	s8	Movie	Sankofa	Haile Gerima	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Kofi Ghanaba	Dramas	Burkina Faso
4	s8	Movie	Sankofa	Haile Gerima	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Kofi Ghanaba	Dramas	United Kingdom

```
final_data.shape
```

```
(3161293, 12)
```

```
final_data.isnull().sum()
```

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

```
final_data.head()
```

	show_id	type	title	director	date_added	release_year	rating	duration	description	cast	listed_in	country
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	September 25, 2021	2020	PG-13	90 min	As her father nears the end of his life, filmm...	nan	Documentaries	United States
1	s8	Movie	Sankofa	Haile Gerima	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Kofi Ghanaba	Dramas	United States
2	s8	Movie	Sankofa	Haile Gerima	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Kofi Ghanaba	Dramas	Ghana
3	s8	Movie	Sankofa	Haile Gerima	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Kofi Ghanaba	Dramas	Burkina Faso
4	s8	Movie	Sankofa	Haile Gerima	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	Kofi Ghanaba	Dramas	United Kingdom

```python
final_data["type"].value_counts()
```

```
Movie      3015506
TV Show     145787
Name: type, dtype: int64
```

```python
final_data.nunique()
```

```
show_id         5747
type               2
title           5747
director        4285
date_added      1516
release_year      73
rating            14
duration         204
description     5736
cast           25459
listed_in         42
country          122
dtype: int64
```
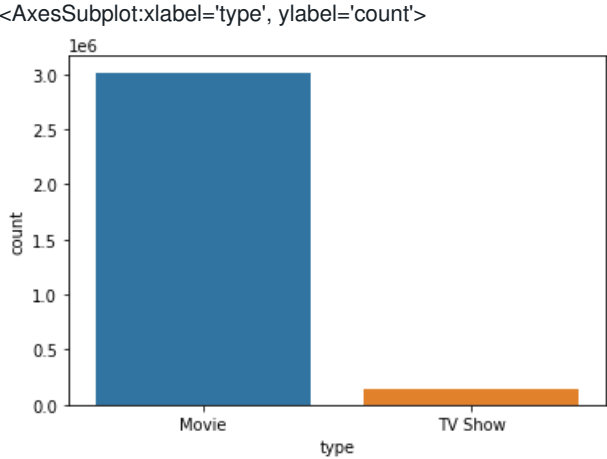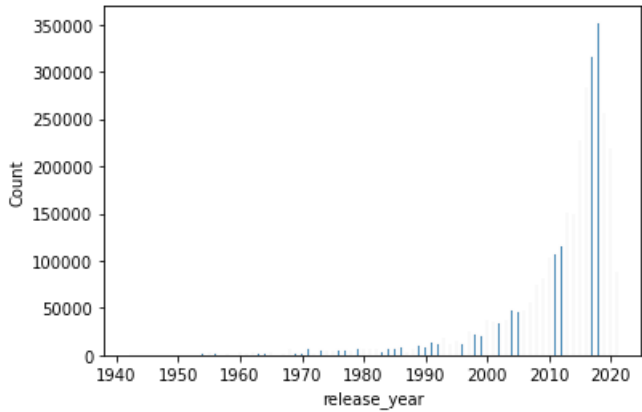
```python
sns.countplot(x="type",data=final_data)
```

```
<AxesSubplot:xlabel='type', ylabel='count'>
```

```python
sns.histplot(data=final_data["release_year"])
```

<AxesSubplot:xlabel='release_year', ylabel='Count'>

```
sns.boxplot(x="type",y="release_year",data=final_data)
```

<AxesSubplot:xlabel='type', ylabel='release_year'>
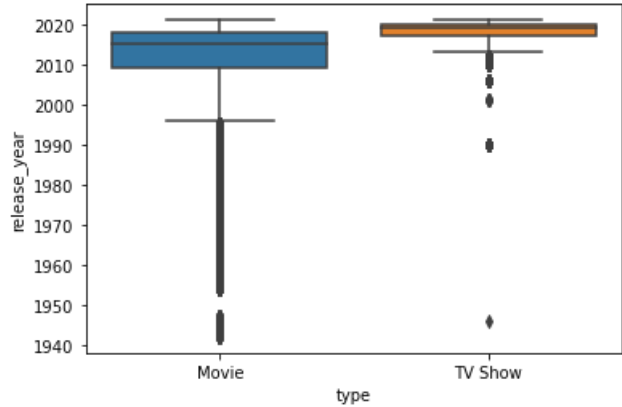
```
sns.pairplot(data=final_data)
```

<seaborn.axisgrid.PairGrid at 0x7fc8a87a88e0>
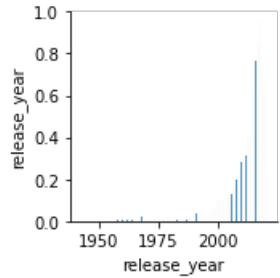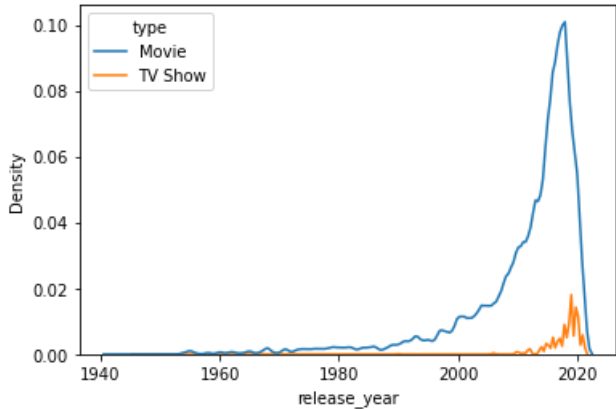
```
sns.kdeplot(x="release_year",hue="type",data=final_data)
```

<AxesSubplot:xlabel='release_year', ylabel='Density'>

```
sns.boxplot(x="rating", y="release_year", data=final_data)
plt.xticks(rotation=90)
```
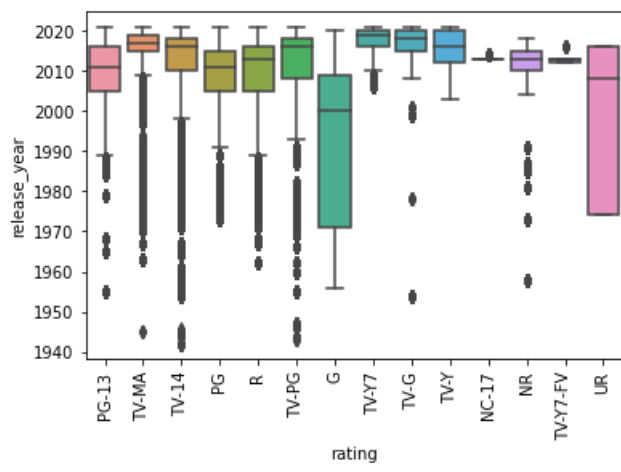
plt.show()

In [ ]: