

In [29]:

```
import pandas as pd  
import numpy as np
```

In [30]:

```
d=pd.read_csv("netflix.csv")
```

In [31]:

d

Out[31]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009

	show_id		type		title	director	cast	country	date_added	release_year
8805	s8806	Movie		Zoom		Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006
8806	s8807	Movie		Zubaan		Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015

8807 rows × 12 columns

In [32]:

```
d.head()
```

Out[32]:

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

In [33]:

d.tail()

Out[33]:

	show_id	type	title	director	cast	country	date_added	release_year	rat
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV

In [34]:

d.shape

Out[34]:

(8807, 12)

In [35]:

d.columns

Out[35]:

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

In [36]:

```
d.dtypes
```

Out[36]:

```
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
```

In [37]:

```
d.describe()
```

Out[37]:

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

In [38]:

```
d.describe(include=['object', 'int'])
```

Out[38]:

	show_id	type	title	director	cast	country	date_added	rating	duration
count	8807	8807	8807	6173	7982	7976	8797	8803	8804
unique	8807	2	8807	4528	7692	748	1767	17	220
top	s8594	Movie	Haven	Rajiv Chilaka	David Attenborough	United States	January 1, 2020	TV-MA	1 Season
freq	1	6131	1	19	19	2818	109	3207	1793

In [39]:

```
d.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
```

Checking duplicated values

In [40]:

```
d.duplicated().sum()
```

Out[40]:

0

Checking Null values

In [41]:

```
d.isna().sum()
```

Out[41]:

```
show_id      0
type         0
title        0
director    2634
cast        825
country     831
date_added   10
release_year  0
rating       4
duration     3
listed_in    0
description  0
dtype: int64
```

In [42]:

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

Out[42]:

```
Rajiv Chilaka      19
Raúl Campos, Jan Suter  18
Marcus Raboy      16
Suhas Kadav       16
Jay Karas         14
..
Tony Datis         1
Michael Barnett    1
Piti Jaturaphat    1
Nacho G. Velilla   1
Castille Landon    1
Name: director, Length: 4528, dtype: int64
```

In [43]:

```
d['country'].value_counts() # Highest number of movies and shows came from USA
```

Out[43]:

```
United States      2818
India              972
United Kingdom     419
Japan              245
South Korea        199
...
United Kingdom, Canada, France, United States  1
Mexico, France, Colombia                      1
Germany, United States, France                 1
Hong Kong, Taiwan                             1
France, Iran, United States                    1
Name: country, Length: 748, dtype: int64
```

In [44]:

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

Out[44]:

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
84 min	1
74 min	1
66 min	1

Name: rating, dtype: int64

1) UnNesting the data

In [45]:

```
d.head()
```

Out[45]:

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

In [106]:

```
data=d.copy()
```

In [107]:

data

Out[107]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009

	show_id	type	title	director	cast	country	date_added	release_year
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015

8807 rows × 12 columns

In [108]:

```
data['director']=data['director'].map(str)
data['cast']=data['cast'].map(str)
data['listed_in']=data['listed_in'].map(str)
```

In [109]:

```
data['director']=data['director'].str.split(',')
data['cast']=data['cast'].str.split(',')
data['listed_in']=data['listed_in'].str.split(',')
```

In [110]:

```
data=data.explode('director')           # Converting all the nested columns into unnested
data=data.explode('cast')
data=data.explode('listed_in')
```

In [111]:

```
data.shape
```

Out[111]:

```
(161216, 12)
```

In [112]:

data

Out[112]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	nan	United States	September 25, 2021	2020	PG-13
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA
1	s2	TV Show	Blood & Water	nan	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA
...
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14

161216 rows × 12 columns



In [113]:

```
g=data.groupby(by='type')
```

In [114]:

```
g
```

Out[114]:

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x000001F70F19E400>
```

In [115]:

```
df1=g.get_group('Movie')
```

In [116]:

df1

Out[116]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	nan	United States	September 25, 2021	2020	F
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	NaN	September 24, 2021	2021	I
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	NaN	September 24, 2021	2021	I
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	James Marsden	NaN	September 24, 2021	2021	I
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Sofia Carson	NaN	September 24, 2021	2021	I
...
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-

111989 rows × 12 columns



In [118]:

```
df2=g.get_group('TV Show')
```

In [119]:

df2

Out[119]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	dis
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	Si
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	Si
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	Si
1	s2	TV Show	Blood & Water	nan	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	Si
1	s2	TV Show	Blood & Water	nan	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	Si
...
8800	s8801	TV Show	Zindagi Gulzar Hai	nan	Hina Khawaja Bayat	Pakistan	December 15, 2016	2012	TV-PG	Si
8800	s8801	TV Show	Zindagi Gulzar Hai	nan	Hina Khawaja Bayat	Pakistan	December 15, 2016	2012	TV-PG	Si
8803	s8804	TV Show	Zombie Dumb	nan	nan	NaN	July 1, 2019	2018	TV-Y7	Si
8803	s8804	TV Show	Zombie Dumb	nan	nan	NaN	July 1, 2019	2018	TV-Y7	Si

	show_id	type	title	director	cast	country	date_added	release_year	rating	di
8803	s8804	TV Show	Zombie Dumb	nan	nan	NaN	July 1, 2019	2018	TV-Y7	S

49227 rows × 12 columns



In []: