

```
In [473...]: # NETFLIX CASE STUDY
```

```
In [ ]: #import necessary libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [ ]:
```

```
In [474...]: # Load the data set into jupyter environment
df=pd.read_csv('netflix.csv')
```

```
In [475...]: #nf_data_attributes
print(f"shape of data frame:{df.shape}")
```



```
shape of data frame:(8807, 12)
```

```
In [476...]: columns=df.columns
print(columns)
```

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

```
In [477...]: # first 5 rows of data frame
df.head(10)
```

	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	50 min
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	50 min
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 hr 1 min
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 hr 1 min
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	50 min
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1 hr 1 min
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	1 hr
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmi Ogunlana, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	1 hr
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins,	United Kingdom	September 24, 2021	2021	TV-14	50 min

show_id	type	title	director	cast	country	date_added	release_year	rating	duration
9	s10	Movie	The Starling	Theodore Melfi	Mary Berry, Paul Ho... Melissa McCarthy, Chris O'Dowd, Kevin Kline, T	United States	September 24, 2021	2021	PG-13

In [478]...

```
"""
=Show_id: Unique ID for every Movie / Tv Show
Type: Identifier - A Movie or TV Show
Title: Title of the Movie / Tv Show
Director: Director of the Movie
Cast: Actors involved in the movie/show-- one or more actors names included
Country: Country where the movie/show was produced-- multiple countries can be listed
Date_added: Date it was added on Netflix
Release_year: Actual Release year of the movie/show
Rating: TV Rating of the movie/show
Duration: Total Duration - in minutes or number of seasons
Listed_in: Genre-- one movie/show related to multiple genre
Description: The summary description
"""

```

Out[478]:

```
'\n>Show_id: Unique ID for every Movie / Tv Show\nType: Identifier - A Movie or TV Show\nTitle: Title of the Movie / Tv Show\nDirector: Director of the Movie\nCast: Actors involved in the movie/show-- one or more actors names included\nCountry: Country where the movie/show was produced-- multiple countries can be listed\nDate_added: Date it was added on Netflix\nRelease_year: Actual Release year of the movie/show\nRating: TV Rating of the movie/show\nDuration: Total Duration - in minutes or number of seasons\nListed_in: Genre-- one movie/show related to multiple genre\nDescription: The summary description\n'
```

In [479]...

```
df.tail()
```

Out[479]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	d
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courtney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	

In [480...]

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
```

In [481]: `df.describe()`

Out[481]: `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 [482]: `df.describe(include="all")`

	show_id	type	title	director	cast	country	date_added	release_year	rating
count	8807	8807	8807	6173	7982	7976	8797	8807.000000	8803
unique	8807	2	8807	4528	7692	748	1767	NaN	17
top	s1	Movie	Dick Johnson Is Dead	Rajiv Chilaka	Attenborough	United States	January 1, 2020	NaN	TV-MA
freq	1	6131	1	19	19	2818	109	NaN	3207
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2014.180198	NaN
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	8.819312	NaN
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1925.000000	NaN
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2013.000000	NaN
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2017.000000	NaN
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2019.000000	NaN
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2021.000000	NaN

In [483]:

```
# unpivot cast names
constraints=df['cast'].apply(lambda x:str(x).split(', ')).tolist()
df_new=pd.DataFrame(constraints,index=df['title'])
df_new=df_new.stack()
df_new=pd.DataFrame(df_new)
df_new.reset_index(inplace=True)
df_new=df_new[['title',0]]
df_new.columns=['title','cast']
df_new
#df_new[df_new['cast']]=='Rupa Bhimani'
```

Out[483]:

	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
...
64946	Zubaan	Manish Chaudhary
64947	Zubaan	Meghna Malik
64948	Zubaan	Malkeet Rauni
64949	Zubaan	Anita Shabdish
64950	Zubaan	Chittaranjan Tripathy

64951 rows × 2 columns

In [484...]

```
#unpivot director names
constraints_d=df['director'].apply(lambda x:str(x).split(', '))
df_newd=pd.DataFrame(constraints_d,index=df['title'])
df_newd=df_newd.stack()
df_newd.reset_index(inplace=True)
df_newd=df_newd[['title',0]]
df_newd.columns=['title','director']
df_newd
```

Out[484]:

	title	director
0	Dick Johnson Is Dead	Kirsten Johnson
1	Blood & Water	nan
2	Ganglands	Julien Leclercq
3	Jailbirds New Orleans	nan
4	Kota Factory	nan
...
9607	Zodiac	David Fincher
9608	Zombie Dumb	nan
9609	Zombieland	Ruben Fleischer
9610	Zoom	Peter Hewitt
9611	Zubaan	Mozez Singh

9612 rows × 2 columns

In [485...]

```
# unpivot country names
constraints_c=df['country'].apply(lambda x:str(x).split(', '))
df_newc=pd.DataFrame(constraints_c,index=df['title'])
df_newc=df_newc.stack()
df_newc.reset_index(inplace=True)
df_newc=df_newc[['title',0]]
df_newc.columns=['title','country']
df_newc
```

Out[485]:

	title	country
0	Dick Johnson Is Dead	United States
1	Blood & Water	South Africa
2	Ganglands	nan
3	Jailbirds New Orleans	nan
4	Kota Factory	India
...
10840	Zodiac	United States
10841	Zombie Dumb	nan
10842	Zombieland	United States
10843	Zoom	United States
10844	Zubaan	India

10845 rows × 2 columns

In [486...]

```
# unpivot listed_in
constraints_g=df['listed_in'].apply(lambda x:str(x).split(', '))
df_newg=pd.DataFrame(constraints_g,index=df['title'])
df_newg=df_newg.stack()
df_newg.reset_index(inplace=True)
df_newg=df_newg[['title',0]]
df_newg.columns=['title','listed_in']
df_newg
```

Out[486]:

	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	Ganglands	Crime TV Shows
...
19318	Zoom	Children & Family Movies
19319	Zoom	Comedies
19320	Zubaan	Dramas
19321	Zubaan	International Movies
19322	Zubaan	Music & Musicals

19323 rows × 2 columns

In [487...]

```
# combine the tables using merge
pda=pd.merge(df_new,df_newd,how='inner',on='title')
pdb=pd.merge(df_newc,df_newg,how='inner',on='title')
pdc=pd.merge(pda,pdb,how='inner',on='title')
pd_t=pd.merge(pdc,df,how='inner',on='title')
pd_t=pd_t.drop(columns=['director_y', 'cast_y','country_y','listed_in_y'])
pd_t.rename(columns = {'cast_x':'cast','director_x':'director','country_x':'country'},inplace=True)
```

Out[487]:

	title	cast	director	country	listed_in	show_id	type	date_added	release_
0	Dick Johnson Is Dead	nan	Kirsten Johnson	United States	Documentaries	s1	Movie	September 25, 2021	
1	Blood & Water	Ama Qamata	nan	South Africa	International TV Shows	s2	TV Show	September 24, 2021	
2	Blood & Water	Ama Qamata	nan	South Africa	TV Dramas	s2	TV Show	September 24, 2021	
3	Blood & Water	Ama Qamata	nan	South Africa	TV Mysteries	s2	TV Show	September 24, 2021	
4	Blood & Water	Khosi Ngema	nan	South Africa	International TV Shows	s2	TV Show	September 24, 2021	
...									
201986	Zubaan	Anita Shabdis	Mozez Singh	India	International Movies	s8807	Movie	March 2, 2019	
201987	Zubaan	Anita Shabdis	Mozez Singh	India	Music & Musicals	s8807	Movie	March 2, 2019	
201988	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	Dramas	s8807	Movie	March 2, 2019	
201989	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	International Movies	s8807	Movie	March 2, 2019	

	title	cast	director	country	listed_in	show_id	type	date_added	release_
201990	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	Music & Musicals	s8807	Movie	March 2, 2019	

201991 rows × 10 columns

In [488]: netflix_data = pd.read_csv('netflix_titles.csv')

In [489]: netflix_data

Out[489]:

	title	cast	director	country	listed_in	show_id	type	date_added	release_
0	Dick Johnson Is Dead	nan	Kirsten Johnson	United States	Documentaries	s1	Movie	September 25, 2021	
1	Blood & Water	Ama Qamata	nan	South Africa	International TV Shows	s2	TV Show	September 24, 2021	
2	Blood & Water	Ama Qamata	nan	South Africa	TV Dramas	s2	TV Show	September 24, 2021	
3	Blood & Water	Ama Qamata	nan	South Africa	TV Mysteries	s2	TV Show	September 24, 2021	
4	Blood & Water	Khosi Ngema	nan	South Africa	International TV Shows	s2	TV Show	September 24, 2021	
...									
201986	Zubaan	Anita Shabdis	Mozez Singh	India	International Movies	s8807	Movie	March 2, 2019	
201987	Zubaan	Anita Shabdis	Mozez Singh	India	Music & Musicals	s8807	Movie	March 2, 2019	
201988	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	Dramas	s8807	Movie	March 2, 2019	
201989	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	International Movies	s8807	Movie	March 2, 2019	

	title	cast	director	country	listed_in	show_id	type	date_added	release_
201990	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	Music & Musicals	s8807	Movie	March 2, 2019	

201991 rows × 10 columns

In [490...]

```
netflix_data[['cast_time', 'type_samp']] = netflix_data.duration.str.split(" ", expand=True)  
netflix_data
```

Out[490]:

	title	cast	director	country	listed_in	show_id	type	date_added	release_
0	Dick Johnson Is Dead	nan	Kirsten Johnson	United States	Documentaries	s1	Movie	September 25, 2021	
1	Blood & Water	Ama Qamata	nan	South Africa	International TV Shows	s2	TV Show	September 24, 2021	
2	Blood & Water	Ama Qamata	nan	South Africa	TV Dramas	s2	TV Show	September 24, 2021	
3	Blood & Water	Ama Qamata	nan	South Africa	TV Mysteries	s2	TV Show	September 24, 2021	
4	Blood & Water	Khosi Ngema	nan	South Africa	International TV Shows	s2	TV Show	September 24, 2021	
...									
201986	Zubaan	Anita Shabdis	Mozez Singh	India	International Movies	s8807	Movie	March 2, 2019	
201987	Zubaan	Anita Shabdis	Mozez Singh	India	Music & Musicals	s8807	Movie	March 2, 2019	
201988	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	Dramas	s8807	Movie	March 2, 2019	
201989	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	International Movies	s8807	Movie	March 2, 2019	

	title	cast	director	country	listed_in	show_id	type	date_added	release_year
201990	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	Music & Musicals	s8807	Movie	March 2, 2019	

201991 rows × 10 columns

In [491]: netflix_data=netflix_data.drop(columns=['duration','type_samp'])

In [492]: netflix_data.head(5)

	title	cast	director	country	listed_in	show_id	type	date_added	release_year	rating
0	Dick Johnson Is Dead	nan	Kirsten Johnson	United States	Documentaries	s1	Movie	September 25, 2021	2020	PG-13
1	Blood & Water	Ama Qamata	nan	South Africa	International TV Shows	s2	TV Show	September 24, 2021	2021	13+
2	Blood & Water	Ama Qamata	nan	South Africa	TV Dramas	s2	TV Show	September 24, 2021	2021	13+
3	Blood & Water	Ama Qamata	nan	South Africa	TV Mysteries	s2	TV Show	September 24, 2021	2021	13+
4	Blood & Water	Khosi Ngema	nan	South Africa	International TV Shows	s2	TV Show	September 24, 2021	2021	13+

In [493]: netflix_data['date_added'] = pd.to_datetime(netflix_data.date_added)
netflix_data.head(30)

Out[493]:

	title	cast	director	country	listed_in	show_id	type	date_added	release_year
0	Dick Johnson Is Dead	nan	Kirsten Johnson	United States	Documentaries	s1	Movie	2021-09-25	2020
1	Blood & Water	Ama Qamata	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021
2	Blood & Water	Ama Qamata	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021
3	Blood & Water	Ama Qamata	nan	South Africa	TV Mysteries	s2	TV Show	2021-09-24	2021
4	Blood & Water	Khosi Ngema	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021
5	Blood & Water	Khosi Ngema	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021
6	Blood & Water	Khosi Ngema	nan	South Africa	TV Mysteries	s2	TV Show	2021-09-24	2021
7	Blood & Water	Gail Mabalane	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021
8	Blood & Water	Gail Mabalane	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021

	title	cast	director	country	listed_in	show_id	type	date_added	release_year
9	Blood & Water	Gail Mabalane	nan	South Africa	TV Mysteries	s2	TV Show	2021-09-24	2021
10	Blood & Water	Thabang Molaba	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021
11	Blood & Water	Thabang Molaba	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021
12	Blood & Water	Thabang Molaba	nan	South Africa	TV Mysteries	s2	TV Show	2021-09-24	2021
13	Blood & Water	Dillon Windvogel	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021
14	Blood & Water	Dillon Windvogel	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021
15	Blood & Water	Dillon Windvogel	nan	South Africa	TV Mysteries	s2	TV Show	2021-09-24	2021
16	Blood & Water	Natasha Thahane	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021
17	Blood & Water	Natasha Thahane	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021

	title	cast	director	country	listed_in	show_id	type	date_added	release_year
18	Blood & Water	Natasha Thahane	nan	South Africa	TV Mysteries	s2	TV Show	2021-09-24	2021
19	Blood & Water	Arno Greeff	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021
20	Blood & Water	Arno Greeff	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021
21	Blood & Water	Arno Greeff	nan	South Africa	TV Mysteries	s2	TV Show	2021-09-24	2021
22	Blood & Water	Xolile Tshabalala	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021
23	Blood & Water	Xolile Tshabalala	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021
24	Blood & Water	Xolile Tshabalala	nan	South Africa	TV Mysteries	s2	TV Show	2021-09-24	2021
25	Blood & Water	Getmore Sithole	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021

	title	cast	director	country	listed_in	show_id	type	date_added	release_year
26	Blood & Water	Getmore Sithole	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021
27	Blood & Water	Getmore Sithole	nan	South Africa	TV Mysteries	s2	TV Show	2021-09-24	2021
28	Blood & Water	Cindy Mahlangu	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021
29	Blood & Water	Cindy Mahlangu	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021

In [540]:

```
netflix_data['cast'].replace('nan','cast not available',inplace=True)
netflix_data['director'].replace('nan','director info n/a',inplace=True)
netflix_data['country'].replace('nan','country info n/a',inplace=True)

netflix_data.head()
```

Out[540]:

	title	cast	director	country	listed_in	show_id	type	date_added	release_year	rat
0	Dick Johnson Is Dead	Dick cast not available	Kirsten Johnson	United States	Documentaries	s1	Movie	2021-09-25	2020	PG
1	Blood & Water	Ama Qamata	director info n/a	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021	
2	Blood & Water	Ama Qamata	director info n/a	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021	
3	Blood & Water	Ama Qamata	director info n/a	South Africa	TV Mysteries	s2	TV Show	2021-09-24	2021	
4	Blood & Water	Khosi Ngema	director info n/a	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021	

In [501...]

```
# directors and film count
directors_movie_count=netflix_data.groupby("director")["title"].nunique()
directors_movie_count=directors_movie_count.sort_values(ascending=False).reset_index()
directors_movie_count.iloc[1:11]
```

Out[501]:

	director	title
1	Rajiv Chilaka	22
2	Jan Suter	21
3	Raúl Campos	19
4	Suhas Kadav	16
5	Marcus Raboy	16
6	Jay Karas	15
7	Cathy Garcia-Molina	13
8	Martin Scorsese	12
9	Youssef Chahine	12
10	Jay Chapman	12

In [512...]

```
director_actor=netflix_data.groupby(["director","cast"])["title"].nunique()
director_actor=director_actor.sort_values(ascending=False).reset_index()
director_actor.loc[director_actor['director']!='director info n/a'].head(50)
```

Out[512]:

	director	cast	title
2	Rajiv Chilaka	Julie Tejwani	19
3	Rajiv Chilaka	Rajesh Kava	19
5	Rajiv Chilaka	Rupa Bhimani	18
6	Rajiv Chilaka	Jigna Bhardwaj	18
9	Rajiv Chilaka	Vatsal Dubey	16
17	Rajiv Chilaka	Mousam	13
20	Rajiv Chilaka	Swapnil	13
68	Suhas Kadav	Saurav Chakraborty	8
84	S.S. Rajamouli	Rana Daggubati	7
85	S.S. Rajamouli	Sathyaraj	7
87	S.S. Rajamouli	Nassar	7
91	Yılmaz Erdoğan	Yılmaz Erdoğan	7
94	S.S. Rajamouli	Prabhas	7
96	S.S. Rajamouli	Anushka Shetty	7
99	S.S. Rajamouli	Tamannaah Bhatia	7
103	S.S. Rajamouli	Ramya Krishnan	7
105	Toshiya Shinohara	Kumiko Watanabe	7
107	Toshiya Shinohara	Satsuki Yukino	7
108	Toshiya Shinohara	Houko Kuwashima	7
111	Toshiya Shinohara	Kappei Yamaguchi	7
121	Toshiya Shinohara	Koji Tsujitani	7
133	Hakan Algül	Ata Demirer	6
149	Joey So	Joseph May	6
152	Omoni Oboli	Omoni Oboli	6
154	Cathy Garcia-Molina	Joross Gamboa	6
155	Tilak Shetty	Smita Malhotra	6
156	David Dhawan	Anupam Kher	6
165	Fernando Ayllón	Ricardo Quevedo	6
179	Quentin Tarantino	Samuel L. Jackson	5
180	Ishi Rudell	Tara Strong	5
181	Ishi Rudell	Ashleigh Ball	5
184	Joey So	Keith Wickham	5
186	Hakan Algül	Salih Kalyon	5
187	Jesse V. Johnson	Scott Adkins	5

	director	cast	title
189	Sooraj R. Barjatya	Salman Khan	5
190	Ishi Rudell	Andrea Libman	5
193	John Paul Tremblay	John Paul Tremblay	5
195	John Paul Tremblay	Mike Smith	5
196	Alex Woo	Kerry Gudjohnsen	5
198	Leslie Small	Kevin Hart	5
199	John Paul Tremblay	Robb Wells	5
206	John Paul Tremblay	Pat Roach	5
208	Alex Woo	Maisie Benson	5
209	Stanley Moore	Paul Killam	5
212	Mike Clattenburg	John Dunsworth	5
215	Mike Clattenburg	John Paul Tremblay	5
216	Ishi Rudell	Rebecca Shoichet	5
217	Sooraj R. Barjatya	Alok Nath	5
220	Robb Wells	Mike Smith	5
223	Robb Wells	Robb Wells	5

In [325]:

```
#cast analysis
cast_count=netflix_data.groupby('cast')['title'].nunique().sort_values(ascending=False)
cast_count.iloc[1:11]
```

Out[325]:

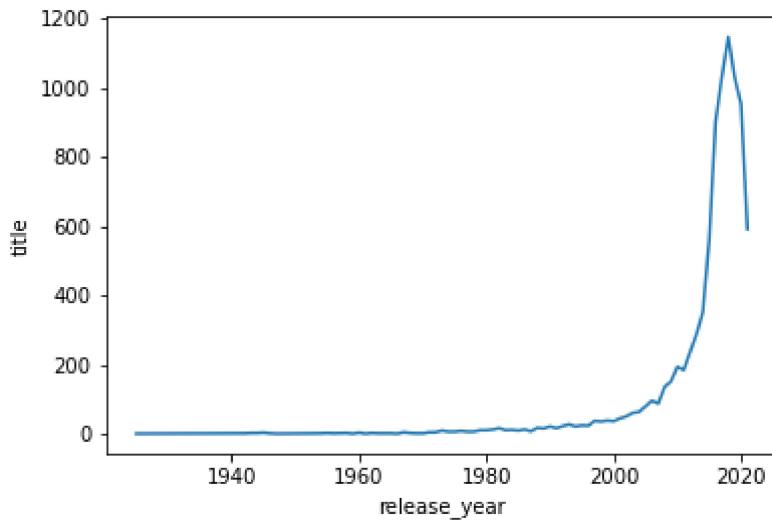
	cast	title
1	Anupam Kher	43
2	Shah Rukh Khan	35
3	Julie Tejwani	33
4	Naseeruddin Shah	32
5	Takahiro Sakurai	32
6	Rupa Bhimani	31
7	Om Puri	30
8	Akshay Kumar	30
9	Yuki Kaji	29
10	Pareesh Rawal	28

In [515]:

```
movies_year=netflix_data.groupby('release_year')['title'].nunique()
movies_year=movies_year.sort_values(ascending=False)
movies_year_last_decade=movies_year.head(10)
```

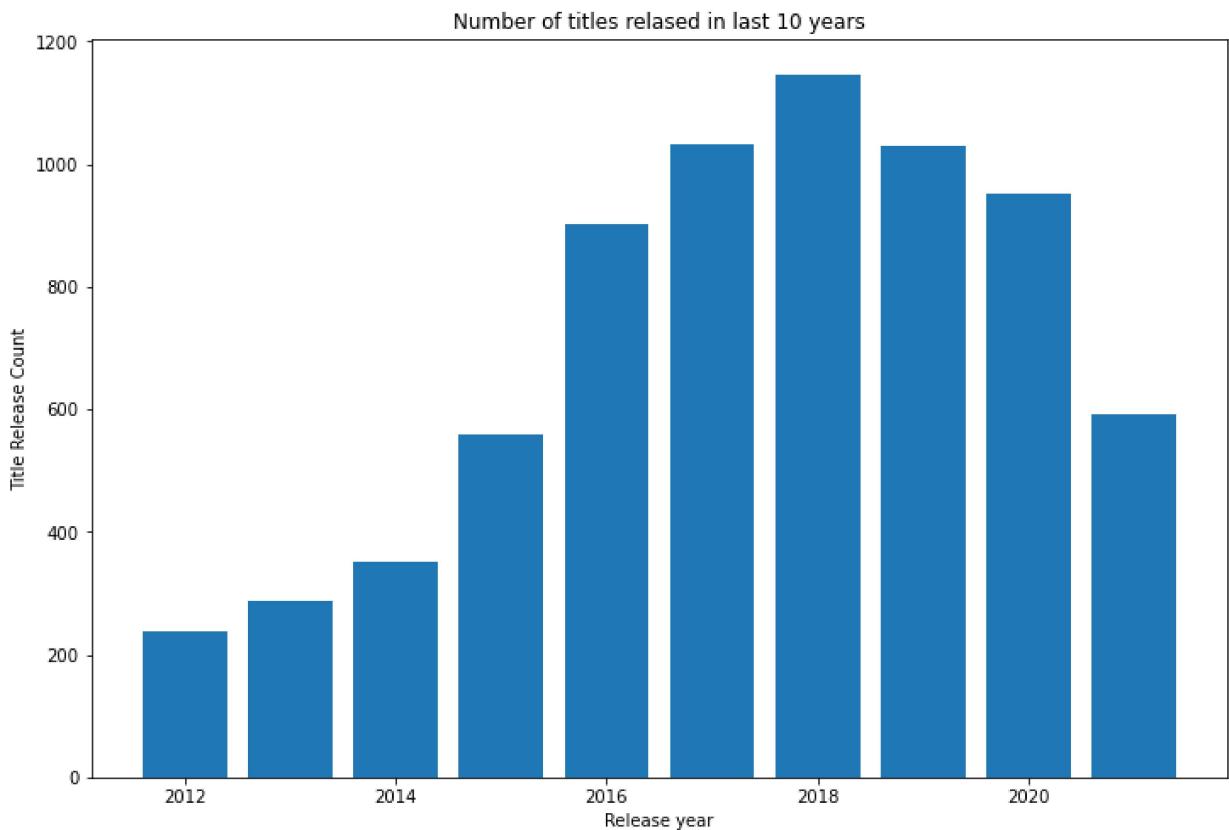
```
In [271]: sns.lineplot(data=movies_year)
```

```
Out[271]: <AxesSubplot:xlabel='release_year', ylabel='title'>
```



```
In [272]: x=movies_year_last_decade.index  
y=movies_year_last_decade  
plt.figure(figsize=(12,8))  
plt.bar(x,y)  
plt.xlabel("Release year")  
plt.ylabel("Title Release Count")  
plt.title("Number of titles released in last 10 years")
```

```
Out[272]: Text(0.5, 1.0, 'Number of titles released in last 10 years')
```



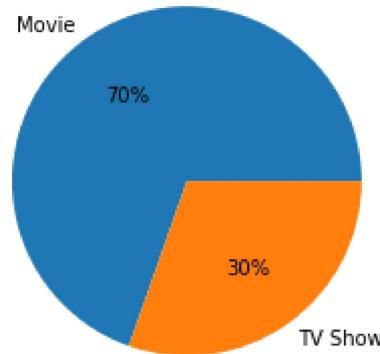
```
In [273]: unique_type_title=netflix_data['type'].unique()
unique_type_title
```

```
Out[273]: array(['Movie', 'TV Show'], dtype=object)
```

```
In [400...]: unique_type_title_count=df.groupby('type')['title'].nunique().reset_index()
unique_type_title_count
```

```
Out[400]:   type  title
0    Movie  6131
1  TV Show  2676
```

```
In [517...]: plt.pie(unique_type_title_count['title'],labels=unique_type_title_count['type'],autopct='%d')
# displaying chart
plt.show()
```



```
In [275...]: # analysing the Listed in genre
genre_count=netflix_data['listed_in'].nunique()
genre_count
```

```
Out[275]: 42
```

```
In [276...]: genre_unique_list=netflix_data['listed_in'].unique()
genre_unique_list
```

```
Out[276]: array(['Documentaries', 'International TV Shows', 'TV Dramas',
       'TV Mysteries', 'Crime TV Shows', 'TV Action & Adventure',
       'Docuseries', 'Reality TV', 'Romantic TV Shows', 'TV Comedies',
       'TV Horror', 'Children & Family Movies', 'Dramas',
       'Independent Movies', 'International Movies', 'British TV Shows',
       'Comedies', 'Spanish-Language TV Shows', 'Thrillers',
       'Romantic Movies', 'Music & Musicals', 'Horror Movies',
       'Sci-Fi & Fantasy', 'TV Thrillers', "Kids' TV",
       'Action & Adventure', 'TV Sci-Fi & Fantasy', 'Classic Movies',
       'Anime Features', 'Sports Movies', 'Anime Series',
       'Korean TV Shows', 'Science & Nature TV', 'Teen TV Shows',
       'Cult Movies', 'TV Shows', 'Faith & Spirituality', 'LGBTQ Movies',
       'Stand-Up Comedy', 'Movies', 'Stand-Up Comedy & Talk Shows',
       'Classic & Cult TV'], dtype=object)
```

```
In [518...]: genre_group_count=netflix_data.groupby('listed_in')[['title']].nunique()  
genre_group_count.sort_values(ascending=False).reset_index()
```

Out[518]:

	listed_in	title
0	International Movies	2752
1	Dramas	2427
2	Comedies	1674
3	International TV Shows	1351
4	Documentaries	869
5	Action & Adventure	859
6	TV Dramas	763
7	Independent Movies	756
8	Children & Family Movies	641
9	Romantic Movies	616
10	TV Comedies	581
11	Thrillers	577
12	Crime TV Shows	470
13	Kids' TV	451
14	Docuseries	395
15	Music & Musicals	375
16	Romantic TV Shows	370
17	Horror Movies	357
18	Stand-Up Comedy	343
19	Reality TV	255
20	British TV Shows	253
21	Sci-Fi & Fantasy	243
22	Sports Movies	219
23	Anime Series	176
24	Spanish-Language TV Shows	174
25	TV Action & Adventure	168
26	Korean TV Shows	151
27	Classic Movies	116
28	LGBTQ Movies	102
29	TV Mysteries	98
30	Science & Nature TV	92
31	TV Sci-Fi & Fantasy	84
32	TV Horror	75
33	Anime Features	71

	listed_in	title
34	Cult Movies	71
35	Teen TV Shows	69
36	Faith & Spirituality	65
37	TV Thrillers	57
38	Movies	57
39	Stand-Up Comedy & Talk Shows	56
40	Classic & Cult TV	28
41	TV Shows	16

```
In [519]: # different ratings
df['rating'].nunique()
```

Out[519]: 17

```
In [527... ratings_data=df.groupby('rating')['title'].nunique().sort_values(ascending=False).reset_index()
ratings_data
ratings_data['percentage']=round((ratings_data['title']/ratings_data['title'].sum())*100,2)
ratings_data
```

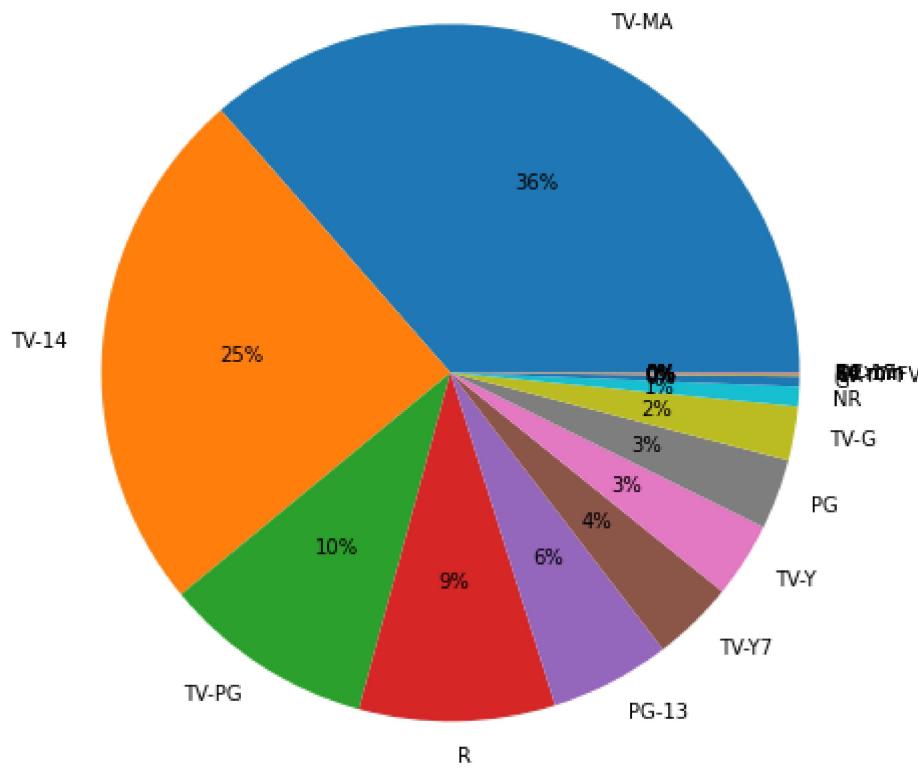
	rating	title	percentage
0	TV-MA	3207	36.43
1	TV-14	2160	24.54
2	TV-PG	863	9.80
3	R	799	9.08
4	PG-13	490	5.57
5	TV-Y7	334	3.79
6	TV-Y	307	3.49
7	PG	287	3.26
8	TV-G	220	2.50
9	NR	80	0.91
10	G	41	0.47
11	TV-Y7-FV	6	0.07
12	UR	3	0.03
13	NC-17	3	0.03
14	74 min	1	0.01
15	84 min	1	0.01
16	66 min	1	0.01

In [539...]

```
plt.figure(figsize=(8,8))

plt.pie(ratings_data['title'], labels=ratings_data['rating'], autopct='%.0f%%')

# displaying chart
plt.show()
```



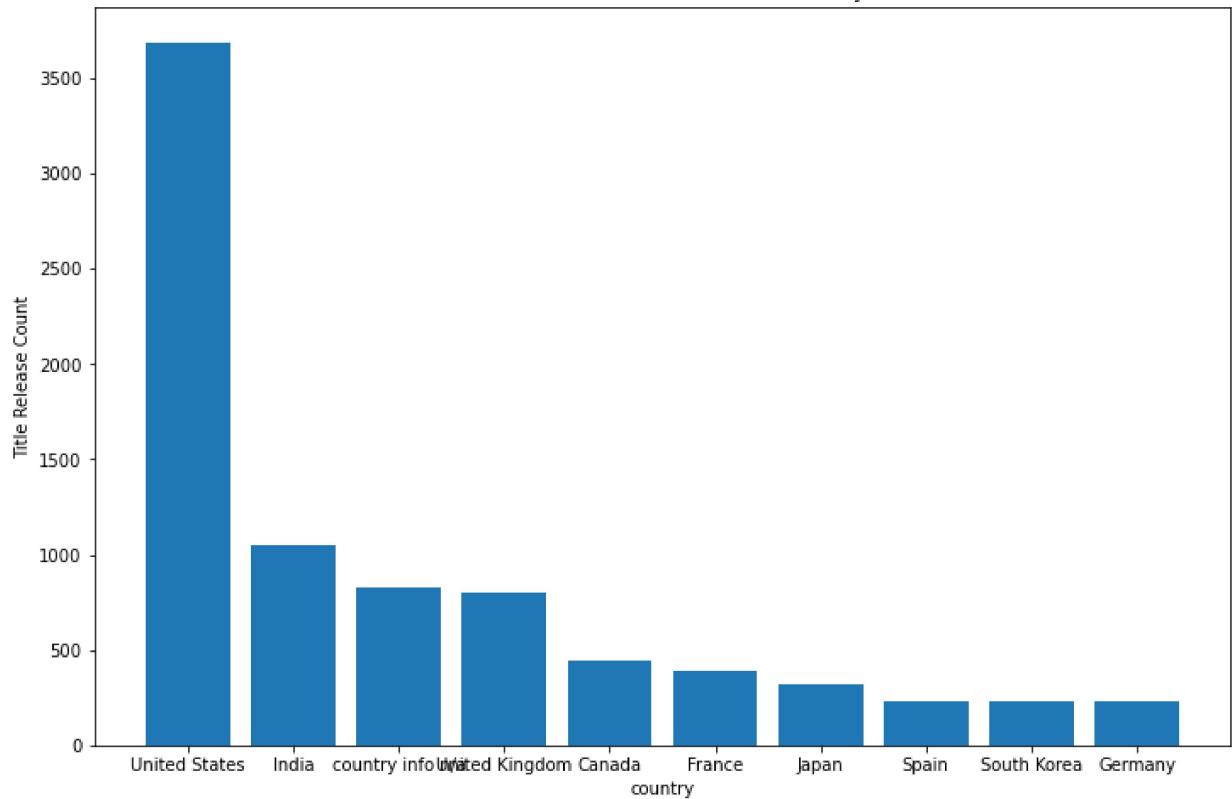
In [550...]

```
data = netflix_data.groupby('country')['title'].nunique().sort_values(ascending=False)
x=data.index
y=data
plt.figure(figsize=(12,8))
plt.bar(x,y)
plt.xlabel("country")
plt.ylabel("Title Release Count")
plt.title("Number of titles released versus country")
```

Out[550]:

```
Text(0.5, 1.0, 'Number of titles released versus country')
```

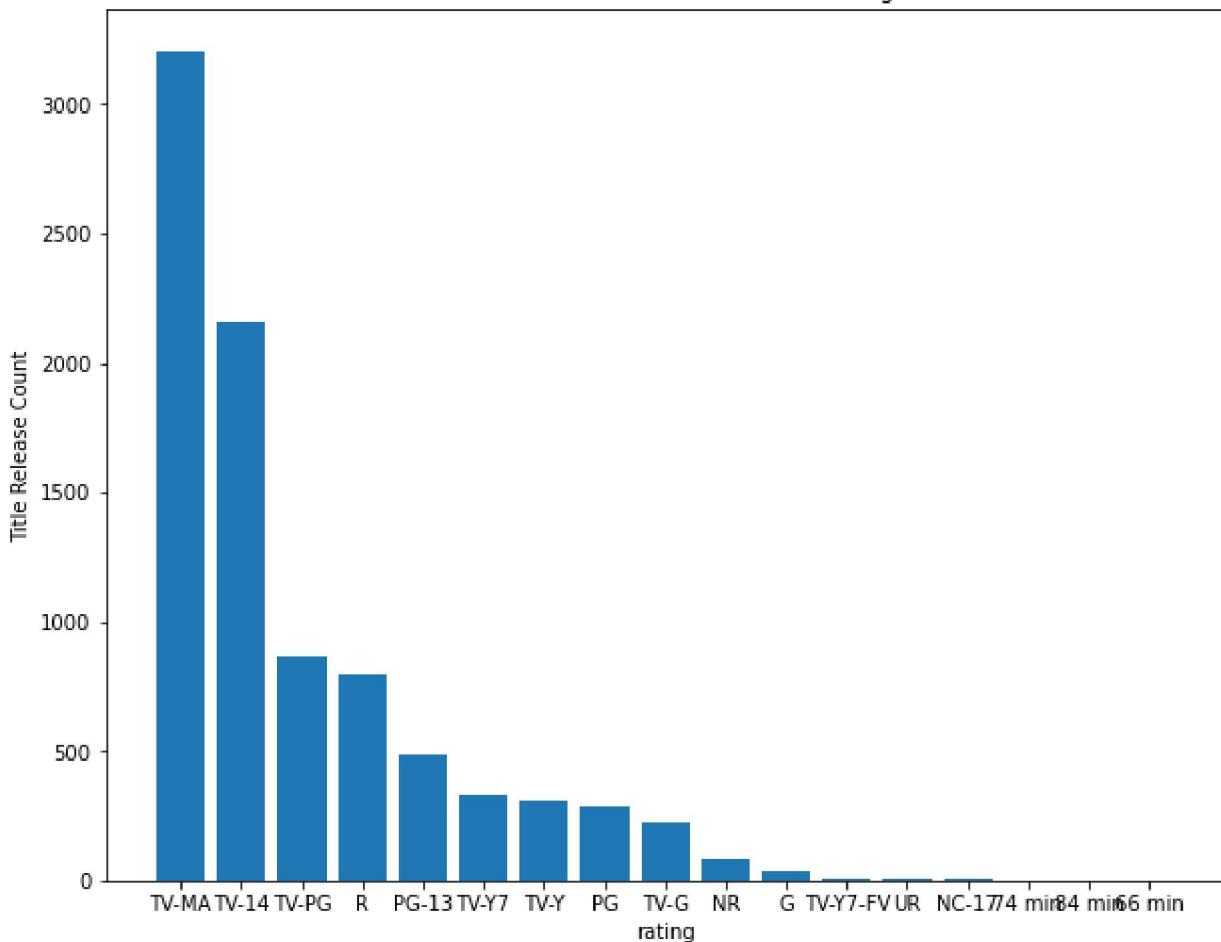
Number of titles released versus country



```
In [543]: rating_order= df.groupby('rating')['title'].count().sort_values(ascending=False)
x=rating_order.index
y=rating_order
plt.figure(figsize=(10,8))
plt.bar(x,y)
plt.xlabel("rating")
plt.ylabel("Title Release Count")
plt.title("Number of titles released versus rating")
```

```
Out[543]: Text(0.5, 1.0, 'Number of titles released versus rating')
```

Number of titles released versus rating



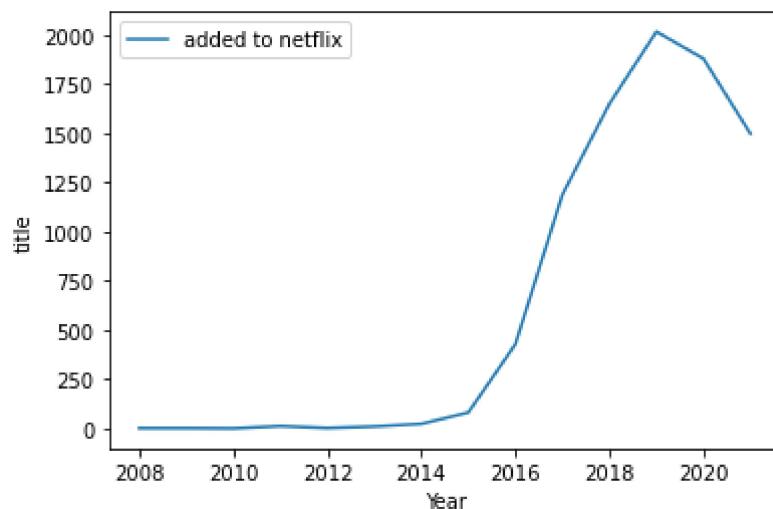
```
In [297]: netflix_data['Year']=netflix_data['date_added'].dt.year
data2=netflix_data.groupby('Year')['title'].nunique().reset_index()
data2.sort_values('title',ascending=False,inplace=True)
data2.head(5)
#data2=netflix_data.groupby('year')['title'].nunique().reset_index()
```

Out[297]:

Year	title
11	2019.0 2016
12	2020.0 1879
10	2018.0 1649
13	2021.0 1498
9	2017.0 1188

```
In [420]: sns.lineplot(data=data2,
                    x="Year",
                    y="title",
                    label="added to netflix")
```

Out[420]: <AxesSubplot:xlabel='Year', ylabel='title'>



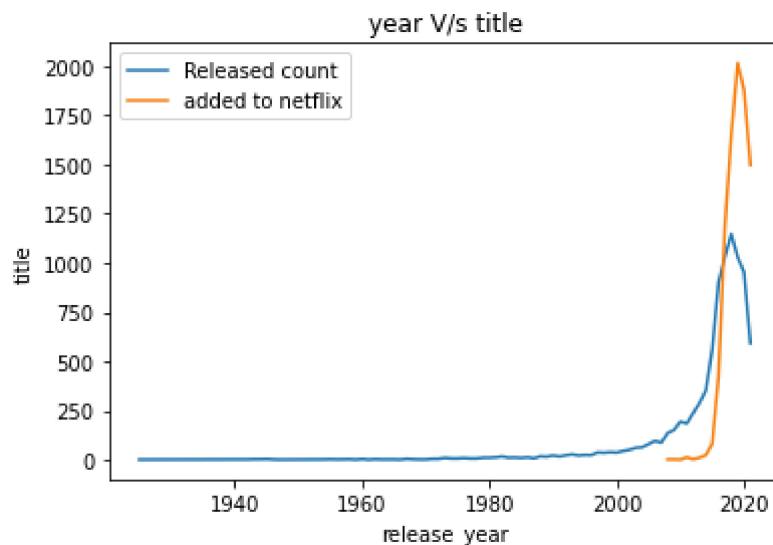
```
In [419...]: movies_release_year=movies_year.reset_index()

sns.lineplot(data=movies_release_year,
              x="release_year",
              y="title",
              label="Released count")

sns.lineplot(data=data2,
              x="Year",
              y="title",
              label="added to netflix")

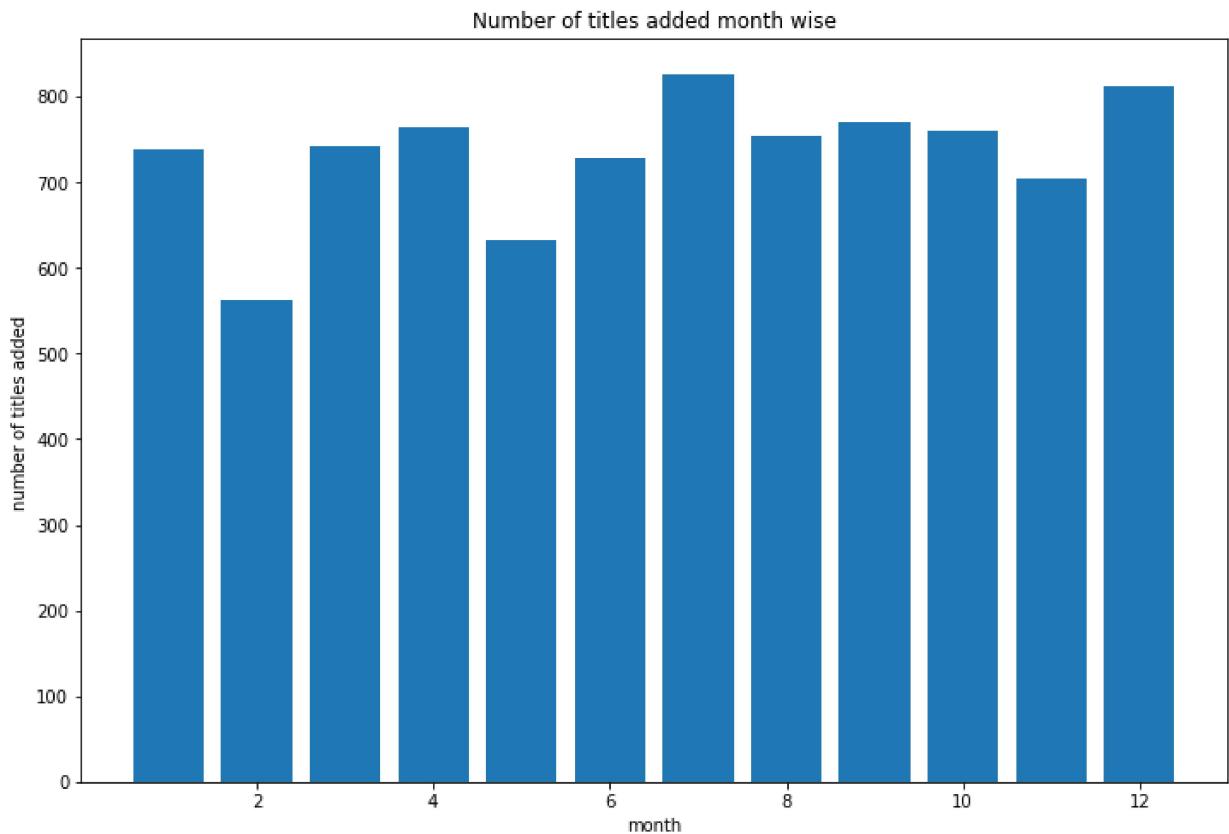
plt.title("year V/s title")
```

Out[419]: Text(0.5, 1.0, 'year V/s title')



```
In [298...]: netflix_data['Month']=netflix_data['date_added'].dt.month
data_month=netflix_data.groupby('Month')['title'].nunique()
x=data_month.index
y=data_month
plt.figure(figsize=(12,8))
plt.bar(x,y)
plt.xlabel("month")
plt.ylabel("number of titles added ")
plt.title("Number of titles added month wise")
```

```
Out[298]: Text(0.5, 1.0, 'Number of titles added month wise')
```



```
In [308...]
```

```
country_type= df[['type', 'country']]
def Mov(x):
    if x == "Movie":
        return 1
def Tv(x):
    if x == "TV Show":
        return 1

country_type['movie'] = list(map(Mov, country_type.type))
country_type['tvshow'] = list(map(Tv, country_type.type))
country_type.drop('type', axis=1, inplace=True)
country_type
```

```
C:\Users\madhu\AppData\Local\Temp\ipykernel_8\887548835.py:9: 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-a-copy
    country_type['movie'] = list(map(Mov, country_type.type))
C:\Users\madhu\AppData\Local\Temp\ipykernel_8\887548835.py:10: 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-a-copy
    country_type['tvshow'] = list(map(Tv, country_type.type))
C:\Users\madhu\AppData\Local\Temp\ipykernel_8\887548835.py:11: 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
    country_type.drop('type', axis=1, inplace=True)
```

Out[308]:

	country	movie	tvshow
0	United States	1.0	NaN
1	South Africa	NaN	1.0
2	NaN	NaN	1.0
3	NaN	NaN	1.0
4	India	NaN	1.0
...
8802	United States	1.0	NaN
8803	NaN	NaN	1.0
8804	United States	1.0	NaN
8805	United States	1.0	NaN
8806	India	1.0	NaN

8807 rows × 3 columns

In [310...]

```
country_type = country_type.groupby('country').sum().sort_values(by='movie', ascending=False)
```

Out[310]:

	movie	tvshow
country		
United States	2058.0	760.0
India	893.0	79.0
United Kingdom	206.0	213.0
Canada	122.0	59.0
Spain	97.0	48.0
Egypt	92.0	14.0
Nigeria	86.0	9.0
Indonesia	77.0	2.0
Japan	76.0	169.0
Turkey	76.0	29.0

In [317...]

```
country_type['movie'] = (country_type['movie']*100)/(country_type['movie'] + country_type['tvshow'])
country_type['tvshow'] =(country_type['tvshow']*100)/(country_type['movie'] + country_type['tvshow'])
```

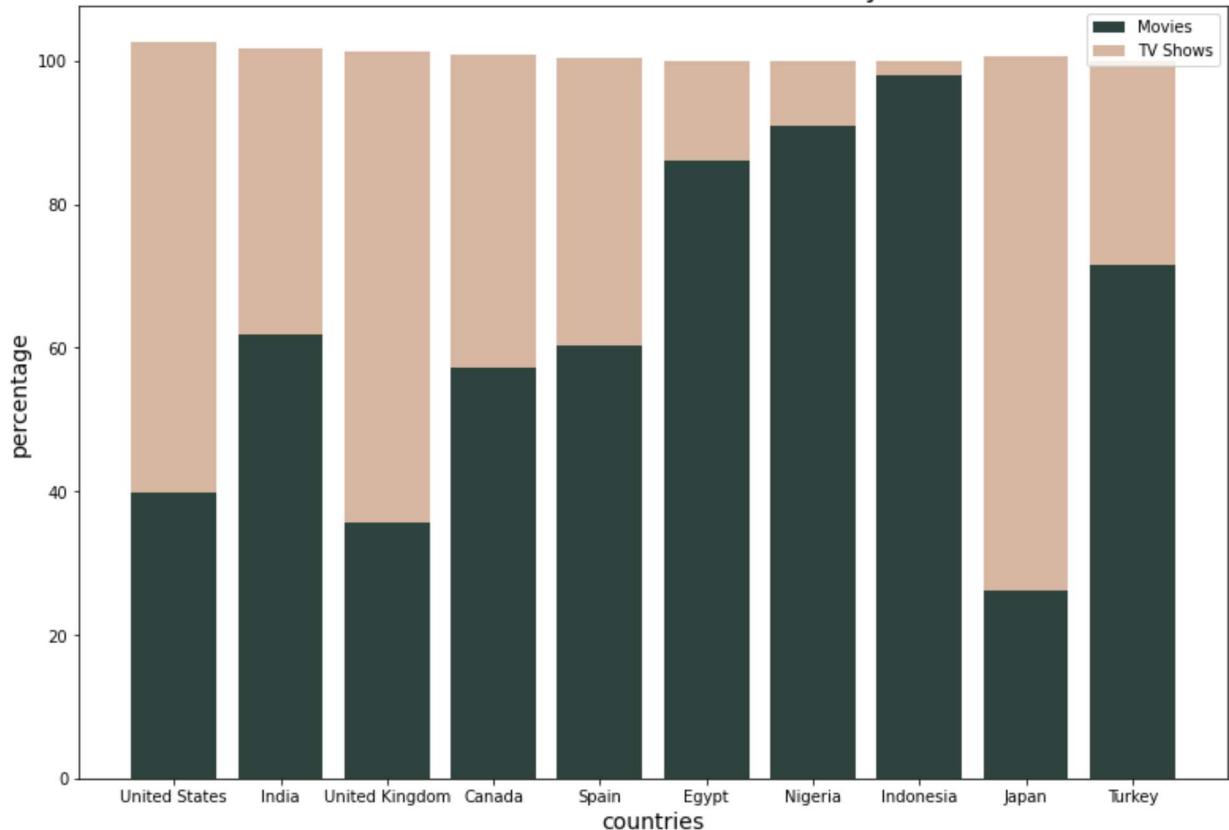
Out[317]:

	movie	tvshow
country		
United States	39.804379	62.813412
India	61.867114	39.858870
United Kingdom	35.564082	65.760761
Canada	57.242414	43.534574
Spain	60.358642	40.112891
Egypt	86.134377	13.875305
Nigeria	90.950405	9.046666
Indonesia	97.988891	2.010906
Japan	26.139452	74.378408
Turkey	71.531510	28.509115

In [318...]

```
plt.figure(figsize=(13, 9))
plt.title("movies and tv shows vs country", fontsize=18)
plt.xlabel("countries", fontsize=14)
plt.ylabel("percentage", fontsize=14)
plt.bar(country_type.index, country_type.movie, color="#2f433e')
plt.bar(country_type.index,country_type.tvshow, bottom=country_type.movie, color='#d7b48c')
plt.legend(['Movies', 'TV Shows'], loc='best')
plt.show()
```

movies and tv shows vs country



In [546...]

```
tv_shows=netflix_data.loc[netflix_data['type']=='TV Show']
tv_shows['num_seasons']=tv_shows['cast_time']
seasons_count=tv_shows.groupby('num_seasons')['title'].nunique().sort_values(ascending=True)
seasons_count.reset_index()
```

C:\Users\madhu\AppData\Local\Temp\ipykernel_8\2264920817.py: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-a-copy

```
tv_shows['num_seasons']=tv_shows['cast_time']
```

Out[546]:

	num_seasons	title
0	1	1793
1	2	425
2	3	199
3	4	95
4	5	65
5	6	33
6	7	23
7	8	17
8	9	9
9	10	7
10	13	3
11	11	2
12	12	2
13	15	2
14	17	1

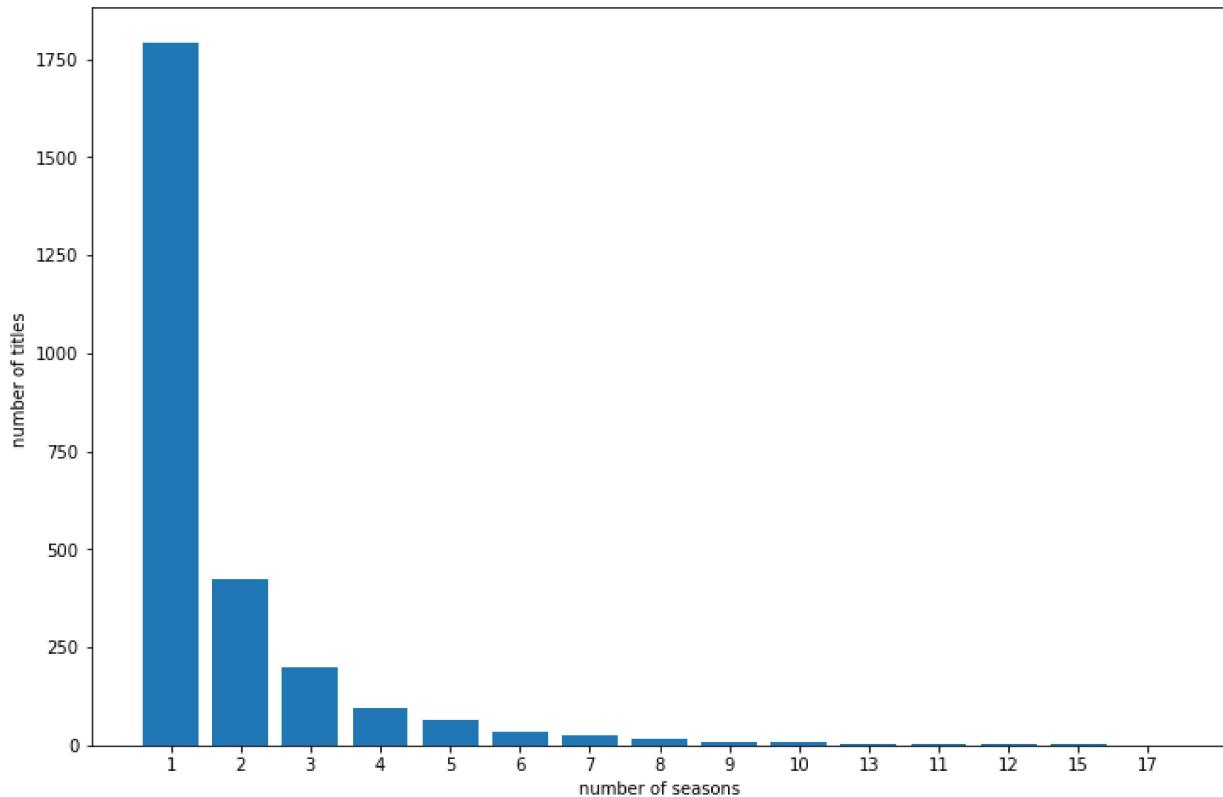
In [365...]

```
x=seasons_count.index
y=seasons_count
plt.figure(figsize=(12,8))
plt.bar(x,y)
plt.xlabel("number of seasons")
plt.ylabel("number of titles ")
plt.title("Number of titles versus seasons")
```

Out[365]:

```
Text(0.5, 1.0, 'Number of titles versus seasons')
```

Number of titles versus seasons



```
In [393]: tv_shows=tv_shows.drop_duplicates(['title','num_seasons'])
tv_shows['num_seasons'] = pd.to_numeric(tv_shows['num_seasons'])
print(f"mean running seasons of Tv shows on netflix: {round(tv_shows['num_seasons'].mean(),2)} seasons")
```

mean running seasons of Tv shows on netflix: 1.76 seasons

```
In [545]: movies=netflix_data.loc[netflix_data['type']=='Movie']
movies_count=movies.groupby('cast_time')[['title']].nunique()
movies_count.sort_values(ascending=False).head(10).reset_index()
```

Out[545]:

	cast_time	title
0	90	152
1	93	146
2	94	146
3	97	146
4	91	144
5	95	137
6	96	130
7	92	129
8	102	122
9	98	120

```
In [389]: movies_unique=movies.drop_duplicates(['title','cast_time'])
print(movies_unique['cast_time'].dtype)
```

```
movies_unique['cast_time'] = pd.to_numeric(movies_unique['cast_time'])
print(f"mean running time of movie on netflix: {round(movies_unique['cast_time'].mean(), 2)} minutes")
object
mean running time of movie on netflix: 99.58 minutes
C:\Users\madhu\AppData\Local\Temp\ipykernel_8\849079772.py:3: 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-a-copy
    movies_unique['cast_time'] = pd.to_numeric(movies_unique['cast_time'])
```

In [394...]: `netflix_data.head()`

	title	cast	director	country	listed_in	show_id	type	date_added	release_year	rating
0	Dick Johnson Is Dead	nan	Kirsten Johnson	United States	Documentaries	s1	Movie	2021-09-25	2020	PG-13
1	Blood & Water	Ama Qamata	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021	14+
2	Blood & Water	Ama Qamata	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24	2021	14+
3	Blood & Water	Ama Qamata	nan	South Africa	TV Mysteries	s2	TV Show	2021-09-24	2021	14+
4	Blood & Water	Khosi Ngema	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24	2021	14+

In [413...]: `netflix_data['listed_in'].nunique()`

Out[413]: 42

```
movies_forecasting=movies_unique.groupby('Year')[['title']].count().sort_values(ascending=True)
movies_forecasting=movies_forecasting.reset_index()
```

```
tv_shows_forecasting=tv_shows.groupby('Year')['title'].count().sort_values(ascending=False)
tv_shows_forecasting=tv_shows_forecasting.reset_index()
```

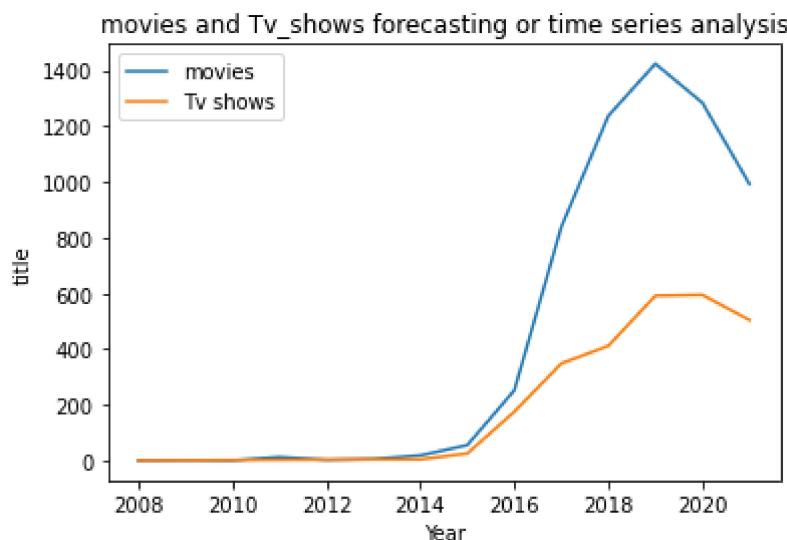
In [433...]

```
sns.lineplot(data=movies_forecasting,
              x="Year",
              y="title",
              label="movies")

sns.lineplot(data=tv_shows_forecasting,
              x="Year",
              y="title",
              label="Tv shows")

plt.title("movies and Tv_shows forecasting or time series analysis")
```

Out[433]:



In [460...]

```
netflix_data.head()
```

Out[460]:

	title	cast	director	country	listed_in	show_id	type	date_added	release_year	ratio
0	Dick Johnson Is Dead	nan	Kirsten Johnson	United States	Documentaries	s1	Movie	2021-09-25 00:00:00	2020	PG-13
1	Blood & Water	Ama Qamata	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24 00:00:00	2021	1.1
2	Blood & Water	Ama Qamata	nan	South Africa	TV Dramas	s2	TV Show	2021-09-24 00:00:00	2021	1.1
3	Blood & Water	Ama Qamata	nan	South Africa	TV Mysteries	s2	TV Show	2021-09-24 00:00:00	2021	1.1
4	Blood & Water	Khosi Ngema	nan	South Africa	International TV Shows	s2	TV Show	2021-09-24 00:00:00	2021	1.1

In [465...]

```
netflix_data['cast'].replace('nan','cast not available',inplace=True)
netflix_data['director'].replace('nan','director info n/a',inplace=True)
```

In [466...]

```
netflix_data.head(5)
```

Out[466]:

	title	cast	director	country	listed_in	show_id	type	date_added	release_year	rat
0	Dick Johnson Is Dead	Dick cast not available	Kirsten Johnson	United States	Documentaries	s1	Movie	2021-09-25 00:00:00	2020	PG
1	Blood & Water	Ama Qamata	director info n/a	South Africa	International TV Shows	s2	TV Show	2021-09-24 00:00:00	2021	
2	Blood & Water	Ama Qamata	director info n/a	South Africa	TV Dramas	s2	TV Show	2021-09-24 00:00:00	2021	
3	Blood & Water	Ama Qamata	director info n/a	South Africa	TV Mysteries	s2	TV Show	2021-09-24 00:00:00	2021	
4	Blood & Water	Khosi Ngema	director info n/a	South Africa	International TV Shows	s2	TV Show	2021-09-24 00:00:00	2021	

In []: