

```
In [129]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
In [5]: import matplotlib.pyplot as plt  
import seaborn as sns
```

```
In [6]: data = pd.read_csv('/Users/divyabansal/Downloads/netflix.csv')
```

```
In [ ]: # Explanatory Analysis
```

```
In [7]: data.head(50)
```

Out [7]:

	show_id	type	title	director	cast	country	date_added
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	NaN	September 24, 2021
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...	United Kingdom	September 24, 2021
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021
10	s11	TV Show	Vendetta: Truth, Lies	NaN	NaN	NaN	September 24, 2021

show_id	type		title	director	cast	country	date_added
			and The Mafia				
11	s12	TV Show	Bangkok Breaking	Kongkiat Komesiri	Sukollawat Kanarot, Sushar Manaying, Pavarit M...	NaN	September 23, 2021
12	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021
13	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Klara Castanho, Lucca Picon, Júlia Gomes, Marc...	NaN	September 22, 2021
14	s15	TV Show	Crime Stories: India Detectives	NaN	NaN	NaN	September 22, 2021
15	s16	TV Show	Dear White People	NaN	Logan Browning, Brandon P. Bell, DeRon Horton,...	United States	September 22, 2021
16	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	Pedro de Echave García, Pablo Azorín Williams	NaN	NaN	September 22, 2021
17	s18	TV Show	Falsa identidad	NaN	Luis Ernesto Franco, Camila Sodi, Sergio Goyri...	Mexico	September 22, 2021
18	s19	Movie	Intrusion	Adam Salky	Freida Pinto, Logan Marshall-Green, Robert Joh...	NaN	September 22, 2021
19	s20	TV Show	Jaguar	NaN	Blanca Suárez, Iván Marcos, Óscar Casas, Adriá...	NaN	September 22, 2021
20	s21	TV Show	Monsters Inside: The 24 Faces of Billy Milligan	Olivier Megaton	NaN	NaN	September 22, 2021

show_id		type	title	director	cast	country	date_added
21	s22	TV Show	Resurrection: Ertugrul	NaN	Engin Altan Düzyatan, Serdar Gökhan, Hülya Dar...	Turkey	September 22, 2021
22	s23	Movie	Avvai Shanmughi	K.S. Ravikumar	Kamal Hassan, Meena, Gemini Ganesan, Heera Raj...	NaN	September 21, 2021
23	s24	Movie	Go! Go! Cory Carson: Chrissy Takes the Wheel	Alex Woo, Stanley Moore	Maisie Benson, Paul Killam, Kerry Gudjohnsen, ...	NaN	September 21, 2021
24	s25	Movie	Jeans	S. Shankar	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...	India	September 21, 2021
25	s26	TV Show	Love on the Spectrum	NaN	Brooke Satchwell	Australia	September 21, 2021
26	s27	Movie	Minsara Kanavu	Rajiv Menon	Arvind Swamy, Kajol, Prabhu Deva, Nassar, S.P....	NaN	September 21, 2021
27	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	September 20, 2021
28	s29	Movie	Dark Skies	Scott Stewart	Keri Russell, Josh Hamilton, J.K. Simmons, Dak...	United States	September 19, 2021
29	s30	Movie	Paranoia	Robert Luketic	Liam Hemsworth, Gary Oldman, Amber Heard, Harr...	United States, India, France	September 19, 2021
30	s31	Movie	Ankahi Kahaniya	Ashwiny Iyer Tiwari, Abhishek Chaubey, Saket C...	Abhishek Banerjee, Rinku Rajguru, Delzad Hiwal...	NaN	September 17, 2021

	show_id	type	title	director	cast	country	date_added
31	s32	TV Show	Chicago Party Aunt	NaN	Lauren Ash, Rory O'Malley, RuPaul Charles, Jil...	NaN	September 17, 2021
32	s33	TV Show	Sex Education	NaN	Asa Butterfield, Gillian Anderson, Ncuti Gatwa...	United Kingdom	September 17, 2021
33	s34	TV Show	Squid Game	NaN	Lee Jung-jae, Park Hae-soo, Wi Ha-jun, Oh Youn...	NaN	September 17, 2021
34	s35	TV Show	Tayo and Little Wizards	NaN	Dami Lee, Jason Lee, Bommie Catherine Han, Jen...	NaN	September 17, 2021
35	s36	Movie	The Father Who Moves Mountains	Daniel Sandu	Adrian Titieni, Elena Porea, Judith State, Val...	NaN	September 17, 2021
36	s37	Movie	The Stronghold	Cédric Jimenez	Gilles Lellouche, Karim Leklou, François Civil...	NaN	September 17, 2021
37	s38	TV Show	Angry Birds	NaN	Antti Pääkkönen, Heljä Heikkinen, Lynne Guagli...	Finland	September 16, 2021
38	s39	Movie	Birth of the Dragon	George Nolfi	Billy Magnussen, Ron Yuan, Qu Jingjing, Terry ...	China, Canada, United States	September 16, 2021
39	s40	TV Show	Chhota Bheem	NaN	Vatsal Dubey, Julie Teiwani, Rupa Bhimani, Jig...	India	September 16, 2021
40	s41	TV Show	He-Man and the Masters of the Universe	NaN	Yuri Lowenthal, Kimberly Brooks, Antony Del Ri...	United States	September 16, 2021

	show_id	type	title	director	cast	country	date_added
41	s42	Movie	Jaws	Steven Spielberg	Roy Scheider, Robert Shaw, Richard Dreyfuss, L...	United States	September 16, 2021
42	s43	Movie	Jaws 2	Jeannot Szwarc	Roy Scheider, Lorraine Gary, Murray Hamilton, ...	United States	September 16, 2021
43	s44	Movie	Jaws 3	Joe Alves	Dennis Quaid, Bess Armstrong, Simon MacCorkind...	United States	September 16, 2021
44	s45	Movie	Jaws: The Revenge	Joseph Sargent	Lorraine Gary, Lance Guest, Mario Van Peebles,...	United States	September 16, 2021
45	s46	Movie	My Heroes Were Cowboys	Tyler Greco	NaN	NaN	September 16, 2021
46	s47	Movie	Safe House	Daniel Espinosa	Denzel Washington, Ryan Reynolds, Vera Farmiga...	South Africa, United States, Japan	September 16, 2021
47	s48	TV Show	The Smart Money Woman	Bunmi Ajakaiye	Osas Ighodaro, Ini Dima-Okojie, Kemi Lala Akin...	NaN	September 16, 2021
48	s49	Movie	Training Day	Antoine Fuqua	Denzel Washington, Ethan Hawke, Scott Glenn, T...	United States	September 16, 2021
49	s50	TV Show	Castle and Castle	NaN	Richard Mofe-Damijo, Dakore Akande, Bimbo Manu...	Nigeria	September 15, 2021

```
In [8]: data.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 [9]: data.shape
```

```
Out[9]: (8807, 12)
```

```
In [10]: data.isnull().sum(axis=0)
```

```
Out[10]: 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 [11]: data.describe()
```

```
Out[11]:
```

	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 [12]: data.describe(include='object')
```

Out [12]:

	show_id	type	title	director	cast	country	date_added	rat
count	8807	8807	8807	6173	7982	7976	8797	88
unique	8807	2	8807	4528	7692	748	1767	
top	s1	Movie	Dick Johnson Is Dead	Rajiv Chilaka	David Attenborough	United States	January 1, 2020	-
freq	1	6131	1	19	19	2818	109	32

In [13]:

data.head()

Out [13]:

	show_id	type	title	director	cast	country	date_added	release_y
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	20
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2

In [262...]

data['date_added']= pd.to_datetime(data['date_added'], errors='coerce')

In [123...]

unpacking the csv columns like director, cast, country, listed_in

In [40]:

constraint=data['cast'].apply(lambda x: str(x).split(', ')).tolist()


```
In [58]: df_new=pd.DataFrame(constraint, index=data['title'])
df_new
```

```
Out[58]:
```

	0	1	2	3	4	5	
title							
Dick Johnson Is Dead	nan	None	None	None	None	None	
Blood & Water	Ama Qamata	Khosi Ngema	Gail Mabalane	Thabang Molaba	Dillon Windvogel	Natasha Thahane	
Ganglands	Sami Bouajila	Tracy Gotoas	Samuel Jouy	Nabiha Akkari	Sofia Lesaffre	Salim Kechiouche	Not
Jailbirds New Orleans	nan	None	None	None	None	None	
Kota Factory	Mayur More	Jitendra Kumar	Ranjan Raj	Alam Khan	Ahsaas Channa	Revathi Pillai	Ur
...	
Zodiac	Mark Ruffalo	Jake Gyllenhaal	Robert Downey Jr.	Anthony Edwards	Brian Cox	Elias Koteas	
Zombie Dumb	nan	None	None	None	None	None	
Zombieland	Jesse Eisenberg	Woody Harrelson	Emma Stone	Abigail Breslin	Amber Heard	Bill Murray	De
Zoom	Tim Allen	Courteney Cox	Chevy Chase	Kate Mara	Ryan Newman	Michael Cassidy	
Zubaan	Vicky Kaushal	Sarah-Jane Dias	Raaghav Chanana	Manish Chaudhary	Meghna Malik	Malkeet Rauni	S

8807 rows x 50 columns

```
In [59]: df_cast= df_new.stack()
```

```
In [60]: df_cast
```

```
Out[60]: title
Dick Johnson Is Dead 0 nan
Blood & Water 0 Ama Qamata
1 Khosi Ngema
2 Gail Mabalane
3 Thabang Molaba
...
Zubaan 3 Manish Chaudhary
4 Meghna Malik
5 Malkeet Rauni
6 Anita Shabdish
7 Chittaranjan Tripathy
Length: 64951, dtype: object
```

```
In [61]: type(df_cast)
```

```
Out[61]: pandas.core.series.Series
```

```
In [115... type(df_cast)
```

```
Out[115... pandas.core.frame.DataFrame
```

```
In [66]: df_cast.reset_index(inplace=True)
```

```
In [67]: df_cast
```

```
Out[67]:
```

	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

```
In [91]: df_cast = pd.DataFrame({'title': df_cast['title'], 'cast': df_cast[0]})
df_cast
```

Out [91]:

	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 [68]: `constraint= data['director'].apply(lambda x: str(x).split(', ')).to_list()`In [69]: `constraint`

```
Out[69]: [['Kirsten Johnson'],
          ['nan'],
          ['Julien Leclercq'],
          ['nan'],
          ['nan'],
          ['Mike Flanagan'],
          ['Robert Cullen', 'José Luis Ucha'],
          ['Haile Gerima'],
          ['Andy Devonshire'],
          ['Theodore Melfi'],
          ['nan'],
          ['Kongkiat Komesiri'],
          ['Christian Schwochow'],
          ['Bruno Garotti'],
          ['nan'],
          ['nan'],
          ['Pedro de Echave García', 'Pablo Azorín Williams'],
          ['nan'],
          ['Adam Salky'],
          ['nan'],
          ['Olivier Megaton'],
          ['nan'],
          ['K.S. Ravikumar'],
          ['Alex Woo', 'Stanley Moore'],
          ['S. Shankar'],
          ['nan'],
          ['Rajiv Menon'],
          ['Dennis Dugan'],
          ['Scott Stewart'],
          ['Robert Luketic'],
          ['Ashwiny Iyer Tiwari', 'Abhishek Chaubey', 'Saket Chaudhary'],
          ['nan'],
          ['nan'],
          ['nan'],
          ['nan'],
          ['Daniel Sandu'],
          ['Cédric Jimenez'],
          ['nan'],
          ['George Nolfi'],
          ['nan'],
          ['nan'],
          ['Steven Spielberg'],
          ['Jeannot Szwarc'],
          ['Joe Alves'],
          ['Joseph Sargent'],
          ['Tyler Greco'],
          ['Daniel Espinosa'],
          ['Bunmi Ajakaiye'],
          ['Antoine Fuqua'],
          ['nan'],
          ['nan'],
          ['Toshiya Shinohara'],
          ['Toshiya Shinohara'],
          ['Toshiya Shinohara'],
          ['Toshiya Shinohara'],
          ['nan'],
          ['Masahiko Murata'],
          ['Hajime Kamegaki'],
          ['Masahiko Murata'],
          ['Hajime Kamegaki'],
```

```
['Masahiko Murata'],  
['Hirotsugu Kawasaki'],  
['Toshiyuki Tsuru'],  
['Tensai Okamura'],  
['David Yarovesky'],  
['nan'],  
['nan'],  
['nan'],  
['Hanns-Bruno Kammertöns', 'Vanessa Nöcker', 'Michael Wech'],  
['nan'],  
['nan'],  
['David A. Vargas'],  
['nan'],  
['Kemi Adetiba'],  
['nan'],  
['Ben Simms'],  
['nan'],  
['Prakash Satam'],  
['Delhiprasad Deenadayalan'],  
['Delhiprasad Deenadayalan'],  
['Tomer Eshed'],  
['Cedric Nicolas-Troyan'],  
['nan'],  
['nan'],  
['JJC Skillz', 'Funke Akindele'],  
['nan'],  
['Thomas Sieben'],  
['nan'],  
['Marcus Clarke'],  
['nan'],  
['Alice Waddington'],  
['Mona Achache', 'Patricia Tourancheau'],  
['nan'],  
['Alexis Almström'],  
['Raja Gosnell'],  
['nan'],  
['Stephen Kijak'],  
['nan'],  
['nan'],  
['nan'],  
['nan'],  
['Chapman Way', 'Maclain Way'],  
['Jason Hehir'],  
['Yemi Amodu'],  
['nan'],  
['Lijo Jose Pellissery'],  
['nan'],  
['David de Vos'],  
['nan'],  
['nan'],  
['Luis Alfaro', 'Javier Gómez Santander'],  
['nan'],  
['Sara Colangelo'],  
['Stephen Herek'],  
['Rahul Rawail'],  
['Jane Campion'],  
['Nagesh Kukunoor'],  
['Luke Holland'],  
['Shanker Raman'],  
['JP Habac'],
```

```
['nan'],  
['nan'],  
['Jane Campion'],  
['nan'],  
['nan'],  
['nan'],  
['Vidhu Vinod Chopra'],  
['Mark Rosman'],  
['Gilles Paquet-Brenner'],  
['Lasse Hallström'],  
['Scott Pleydell-Pearce'],  
['Ridley Scott'],  
['nan'],  
['Neill Blomkamp'],  
['Phillip Noyce'],  
['Renny Harlin'],  
['Anthony Minghella'],  
['Simon Wincer'],  
['Lasse Hallström'],  
['Spike Lee'],  
['Sebastián Schindel'],  
['Steven C. Miller'],  
['Richard LaGravenese'],  
['Martin Campbell'],  
['Reginald Hudlin'],  
['George Jackson', 'Doug McHenry'],  
['Eric Meza'],  
['nan'],  
['Gerhard Mostert'],  
['Michael Martin'],  
['Michael Rymer'],  
['Andrew Lau Wai-keung', 'Alan Mak'],  
['Brett Weiner'],  
['nan'],  
['nan'],  
['Jim Henson'],  
['Gary Winick'],  
['Danishka Esterhazy'],  
['Troy Byer'],  
['Pang Ho-cheung'],  
['nan'],  
['Tim Burton'],  
['Reginald Hudlin'],  
['David Zucker'],  
['Kinka Usher'],  
['nan'],  
['Sergio Leone'],  
['Matthew O'Callaghan', 'Todd Wilderman'],  
['Bobby Farrelly', 'Peter Farrelly'],  
['Wolfgang Petersen'],  
['Peter Spierer'],  
['Michael Carney'],  
['Richard Linklater'],  
['Amy Rice'],  
['Antoine Fuqua'],  
['Randal Kleiser'],  
['Michael Ritchie'],  
['J. Lee Thompson'],  
['Evan Goldberg', 'Seth Rogen'],  
['Tom Shadyac'],
```

```
['Peter Segal'],
['nan'],
['Malcolm D. Lee'],
['Wolfgang Petersen'],
['nan'],
['Chapman Way', 'MacLain Way'],
['nan'],
['nan'],
['Ramzy Bedia', 'Éric Judor'],
['nan'],
['Sharan Koppisetty'],
['Taylor Sheridan'],
['Sachin Yardi'],
['nan'],
['nan'],
['Saurabh Kabra'],
['Mark Waters'],
['nan'],
['Kemi Adetiba'],
['Partho Mitra'],
['Santram Varma'],
['Anil V. Kumar', 'Anurag Basu'],
['Sangeeth Sivan'],
['Umesh Ghadge'],
['Sachin Yardi'],
['David Dhawan'],
['Dibakar Banerjee'],
['Apoorva Lakhia'],
['Milan Luthria'],
['Milan Luthria'],
['Pawan Kripalani'],
['Bhushan Patel'],
['nan'],
['nan'],
['Magnus Martens'],
['Apoorva Lakhia'],
['Raj Nidimoru', 'Krishna D.K.'],
['Milan Luthria'],
['nan'],
['nan'],
['nan'],
['Joshua Rofé'],
['Brad Anderson'],
['Mauricio Dias', 'Tatiana Villela'],
['nan'],
['nan'],
['nan'],
['Angel Kristi Williams'],
['Roger Donaldson'],
['Christopher Alender'],
['Rush Sturges'],
['David Oyelowo'],
['nan'],
['Mark Lo'],
['nan'],
['Crystal Moselle'],
['Rathindran R Prasad'],
['Rathindran R Prasad'],
['Rathindran R Prasad'],
['Rathindran R Prasad'],
```

```
['Han Kwang Il'],
['nan'],
['nan'],
['nan'],
['Karim El Shenawy'],
['nan'],
['Yin Chen-hao'],
['Brian Andrew Mendoza'],
['nan'],
['Dave Needham'],
['Veronica Velasco'],
['Drake Doremus'],
['Miguel Alexandre'],
['Mani Ratnam'],
['nan'],
['Michael Harte'],
['Tosin Igho'],
['Mani Ratnam'],
['Alice Filippi'],
['Paakhi Tyrewala'],
['nan'],
['Bruno Garotti'],
['Laura Brownson'],
['nan'],
['Steve Brill'],
['nan'],
['nan'],
['Jane Champion'],
['Moses Inwang'],
['nan'],
['Ferdinando Cito Filomarino'],
['nan'],
['nan'],
['Juan Carlos Medina'],
['nan'],
['nan'],
['Inma Torrente'],
['nan'],
['Julián Gaviria'],
['Steven Yamamoto'],
['nan'],
['Charles Uwagbai'],
['Julián Hernández'],
['Sam Hobkinson'],
['Vince Marcello'],
['Jonathan Teplitzky'],
['nan'],
['Sakon Tiacharoen'],
['nan'],
['nan'],
['Floyd Russ'],
['nan'],
['Dustin Hoffman'],
['Adze Ugah'],
['Hideaki Takizawa'],
['Quoc Bao Tran'],
['nan'],
['Bejoy Nambiar'],
['Priyadarshan'],
['Karthik Narain'],
```


'Vasanth Sai',
'Karthik Subbaraj',
'Arvind Swamy',
'Rathindran R Prasad',
'Sarjun',
'Gautham Vasudev Menon'],
['Kayode Kasum'],
['Just Philippot'],
['Kirk DeMicco', 'Brandon Jeffords'],
['Rohit Shetty'],
['Cavi Borges', 'Luciano Vidigal'],
['Alejandro Doria'],
['Laura Fairrie'],
['Marcelo Piñeyro'],
['Izu Ojukwu'],
['Peter Winther'],
['Susan Lacy'],
['nan'],
['Billy Corben'],
['nan'],
['nan'],
['nan'],
['James Mangold'],
['Chineze Anyaene'],
['Hsu Fu-chun'],
['Kristine Stolakis'],
['Eva Müller', 'Michael Schmitt'],
['nan'],
['nan'],
['nan'],
['nan'],
['nan'],
['nan'],
['Brian Levant'],
['Rod Daniel'],
['Robert Zemeckis'],
['Todor Chapkanov'],
['Steven Spielberg'],
['Phil Lord', 'Christopher Miller'],
['nan'],
['Renny Harlin'],
['John Hughes'],
['Justin Baldoni'],
['Joe Roth'],
['nan'],
['Mark Helfrich'],
['nan'],
['Mag Hsu', 'Hsu Chih-yen'],
['Christopher Nolan'],
['Paul Thomas Anderson'],
['Nick Castle'],
['Howard Zieff'],
['Howard Zieff'],
['David Feiss'],
['David Gordon Green'],
['Jorge Blanco'],
['Zara Hayes'],
['Gary Ross'],
['Clint Eastwood'],
['Trey Parker'],

```
['Kelly Fremon Craig'],
['Tom Elkins'],
['Tony Scott'],
['Brad Furman'],
['Sylvain White'],
['Brad Anderson'],
['Irwin Winkler'],
['Spike Lee'],
['Garry Marshall'],
['nan'],
['Kenneth Gyang'],
['Jaume Balagueró'],
['nan'],
['Najwa Najjar'],
['nan'],
['Selçuk Metin'],
['nan'],
['nan'],
['nan'],
['Najwa Najjar'],
['Keishi Otomo'],
['David Charhon'],
['Michelle Bello'],
['Steven Tsuchida'],
['nan'],
['Daniel Markowicz'],
['Louie Schwartzberg'],
['nan'],
['Glen Winter'],
['nan'],
['nan'],
['Ram Gopal Varma'],
['David Benullo'],
['nan'],
['nan'],
['Laxman Utekar'],
['Royale Watkins', 'Rich Schlansker'],
['Yuval Adler'],
['nan'],
['nan'],
['Quentin Tarantino'],
['Clay Glen'],
['Muzi Mthembu'],
['Marcos Bucay'],
['Peter Thorwarth'],
['nan'],
['Kim Seong-hun'],
['nan'],
['nan'],
['nan'],
['Augustine Frizzell'],
['nan'],
['nan'],
['Sidheswar Shukla'],
['Rajiv Chilaka'],
['Rajiv Chilaka'],
['Rajiv Chilaka'],
['Rajiv Chilaka'],
['Rajiv Chilaka'],
['Rajiv Chilaka'],
```

```
['Rajiv Chilaka'],
['Rajiv Chilaka'],
['Rajiv Chilaka'],
['Rajiv Chilaka'],
['Rajiv Chilaka'],
['Binayak Das'],
['Arpan Sarkar', 'Shyamal Chaulia'],
['Rajiv Chilaka'],
['Rajiv Chilaka'],
['Rajiv Chilaka', 'Owll Mina'],
['Rajiv Chilaka'],
['Rajiv Chilaka'],
['Rajiv Chilaka'],
['nan'],
['Ainsley Gardiner', 'Briar Grace-Smith'],
['Mahmoud Karim'],
['Kyohei Ishiguro'],
['Seyi Babatope'],
['nan'],
['nan'],
['Johane Matte', 'Andrew L. Schmidt', 'Francisco Ruiz Velasco'],
['Morgan Ingari'],
['nan'],
['nan'],
['Abdulaziz Alshlahei'],
['Edward Drake'],
['Kathryn Fasegha'],
['Sita Likitvanichkul',
 'Jetarin Ratanaserikiat',
 'Apirak Samudkitpaisan',
 'Thanabodee Uawithya',
 'Adirek Wattaleela'],
['nan'],
['Leigh Janiak'],
['nan'],
['Luis Estrada'],
['Garrett Bradley'],
['Sofia Coppola'],
['Colin Trevorrow'],
['Bill Condon'],
['Bill Condon'],
['David Slade'],
['Chris Weitz'],
['Catherine Hardwicke'],
['Hadrach Daeng Ratu'],
['nan'],
['Tommy Chong'],
['Fred Ouro Preto'],
['nan'],
['Mayye Zayed'],
['Alessandra de Rossi'],
['nan'],
['Ash Brannon', 'Chris Buck'],
['Alaa Eddine Aljem'],
['Tom Donahue'],
['Roberto De Feo', 'Paolo Strippoli'],
['Navot Papushado'],
['nan'],
['nan'],
['Manuel Alcalá'],
```

[illegible]

```
['Rajveer Singh Maan', 'Harpeet Singh'],  
['Youssef Chahine'],  
['nan'],  
['Saw Teong Hin', 'Nik Amir Mustapha', 'M.S. Prem Nath'],  
['nan'],  
['nan'],  
['Rano Karno'],  
['nan'],  
['nan'],  
['nan'],  
['nan'],  
['nan'],  
['nan'],  
['Pramod Pawar'],  
['Naresh Saigal'],  
['nan'],  
['nan'],  
['nan'],  
['nan'],  
['nan'],  
['nan'],  
['nan'],  
['Leigh Janiak'],  
['Vinil Mathew'],  
['Neri Parenti'],  
['nan'],  
['Ramsey Nouah'],  
['Bong Joon Ho'],  
['Kim Tae-hyung'],  
['Michael Spierig', 'Peter Spierig'],  
['Ernie Barbarash'],  
['Wolfgang Petersen'],  
['Matt Ogens'],  
['Jay Roach'],  
['Jay Roach'],  
['Jay Roach'],  
['Paul Thomas Anderson'],  
['nan'],  
['McG'],  
['Frank Marshall'],  
['Nick Castle'],  
['Victor Vu'],  
['Chow Hin Yeung Roy'],  
['nan'],  
['Joel Hopkins'],  
['John Stevenson', 'Mark Osborne'],  
['Jennifer Yuh Nelson'],  
['Greg Berlanti'],  
['Garth Davis'],  
['Malik Nejer'],  
['Rob Marshall'],  
['Martin Brest'],  
['Shuko Murase'],  
['Paul W.S. Anderson'],  
['Garry Marshall'],  
['Ivan Reitman'],  
['Joel Gallen'],  
['Claire McCarthy'],  
['Mary Lambert'],  
['Sidharta Tata'],
```

'Aco Tenriyagelli',
'Dian Sastrowardoyo',
'Ifa Isfansyah',
'Jason Iskandar'],
['nan'],
['Trevor Nunn'],
['nan'],
['Gabriele Muccino'],
['Jim Field Smith'],
['Chris Koch'],
['J.J. Abrams'],
['Rob Minkoff'],
['Lynn Shelton'],
['Adam McKay'],
['Anton Corbijn'],
['Robin Bissell'],
['David Fincher'],
['John G. Avildsen'],
['John G. Avildsen'],
['John G. Avildsen'],
['Alan Parker'],
['Walter Hill'],
['Stephen Frears'],
['Bryan Bertino'],
['Phil Alden Robinson'],
['Florian Henckel von Donnersmarck'],
['Len Wiseman'],
['Måns Mårild', 'Björn Stein'],
['Patrick Tatopoulos'],
['Robert O. Peters'],
['Vincent Ward'],
['Gregory Nava'],
['Mary Harron'],
['nan'],
['Matt Thompson'],
['Jameel Buari'],
['nan'],
['Shen Leping'],
['Matt Aselton'],
['Jose Javier Reyes'],
['Jakub Piątek'],
['nan'],
['John Dower'],
['nan'],
['nan'],
['B. T Thomas'],
['Andrew Dominik'],
['nan'],
['Nibal Arakji'],
['Sofie Šustková'],
['Anissa Bonnefont'],
['Bahij Hojeij'],
['Jonathan Hensleigh'],
['Srijit Mukherji', 'Vasan Bala', 'Abhishek Chaubey'],
['nan'],
['nan'],
['nan'],
['nan'],
['nan'],
['Justin P. Lange'],

```
['Kimmy Gatewood'],
['Tània Balló'],
['Manolo Caro'],
['nan'],
['nan'],
['Anurin Nwunembom'],
['Jayme Monjardim'],
['Kingsley Ogoro'],
['Kingsley Ogoro'],
['nan'],
['Cristina Jacob'],
['Cristina Jacob'],
['Cristina Jacob'],
['nan'],
['nan'],
['nan'],
['nan'],
['nan'],
['nan'],
['nan'],
['nan'],
['Fernando Moro'],
['nan'],
['Ivan Andrew Payawal'],
['nan'],
['nan'],
['nan'],
['Yoshiyuki Tomino', 'Yoshikazu Yasuhiko'],
['Yoshiyuki Tomino', 'Yoshikazu Yasuhiko'],
['Yoshiyuki Tomino'],
['nan'],
['nan'],
['nan'],
['Rob Seidenglanz'],
['nan'],
['nan'],
['nan'],
['nan'],
['nan'],
['Simon Frederick'],
['nan'],
['Michihito Fujii'],
['nan'],
['Paul Weitz'],
['nan'],
['Karthik Subbaraj'],
['Kayode Kasum'],
['Yoshiyuki Tomino', 'Yoshikazu Yasuhiko'],
['Keishi Otomo'],
['nan'],
['Hsu Fu-chun'],
['Lucky Kuswandi'],
['Soudade Kaadan'],
['Soudade Kaadan'],
['nan'],
['nan'],
['Antoinette Jadaone'],
['nan'],
['nan'],
['Marie Clements'],
['David O. Russell'],
['nan'],
```

```
['Achille Brice'],
['Max Jabs'],
['nan'],
['Sarawut Wichiensarn'],
['Ricardo de Montreuil'],
['nan'],
['nan'],
['Peter Chelsom'],
['Michael Lockshin'],
['nan'],
['nan'],
['nan'],
['Francine Parker'],
['nan'],
['Daniel Schechter'],
['nan'],
['nan'],
['Mike Gunther'],
['David Zeiger'],
['Michael Simon'],
['Jerry Rothwell'],
['nan'],
['Joe Berlinger', 'Bruce Sinofsky'],
['Ian Cheney', 'Sharon Shattuck'],
['Bradley Parker'],
['nan'],
['nan'],
['nan'],
['nan'],
['nan'],
['nan'],
['Prabhakaran'],
['Manjari Makijany'],
['Jay Oliva'],
['nan'],
['Chris Appelhans'],
['Michael Tiddes'],
['Lara Saba'],
['Bao Nhan', 'Namcito'],
['Mark Raso'],
['nan'],
['Tariq Alkazim'],
['Mark Raso'],
['Kenneth Gyang'],
['nan'],
['nan'],
['Yulene Olaizola'],
['nan'],
['Mark Waters'],
['nan'],
['Matthew Heineman'],
['nan'],
['Robert Peters'],
['Jonathan Clay'],
['Ally Pankiw'],
['George Ford'],
['George Ford'],
['nan'],
['Lee Kae-byeok'],
['nan'],
```



```
['Jayan Moodley'],
['Daniel Benmayor'],
['Alex Díaz'],
['nan'],
['Helena Bergström'],
['Alejandro De Grazia', 'Juan Stadler'],
['Myriam Fares'],
['Chiaki Kon'],
['nan'],
['Rae Red'],
['Lance Hool'],
['Nicole Conn'],
['Les Mayfield'],
['Ellen Seidler', 'Megan Siler'],
['nan'],
['Roger Christian'],
['nan'],
['Peter Galison'],
['Lance Young'],
['Leandro Neri'],
['Thom Fitzgerald'],
['nan'],
['Don Michael Paul'],
['Andrzej Bartkowiak'],
['Harold Becker'],
['nan'],
['Andrew Jenks'],
['Ric Roman Waugh'],
['Rob Reiner'],
['Andy Tennant'],
['Tade Ogidan'],
['nan'],
['Scott Spiegel'],
['Jessie Nelson'],
['Theodore Witcher'],
['Clint Eastwood'],
['Aurora Guerrero'],
['Michael Jai White'],
['James McTeigue'],
['Takuya Igarashi'],
['Michelle MacLaren'],
['nan'],
['Hidenori Inoue'],
['Hidenori Inoue'],
['Don Michael Paul'],
['Todd Phillips'],
['Walter Hill'],
['Steve Ball'],
['Dominic Sena'],
['nan'],
['Alan Alda'],
['Mika Kaurismäki'],
['Sydney Pollack'],
['Lorene Scafaria'],
['Barbra Streisand'],
['Clint Eastwood'],
['nan'],
['Michael Winterbottom'],
['Emma Tammi'],
['Peter Hutchings'],
```

```
['George Ratliff'],
['nan'],
['Bo Burnham'],
['Ekene Som Mekwunye'],
['David Frankel'],
['Kevin Johnson'],
['nan'],
['José Larraza', 'Marc Pons'],
['Gary Sing'],
['nan'],
['Julio Quintana'],
['nan'],
['Paween Purijitpanya'],
['nan'],
['Pablo Faro'],
['Soudade Kaadan'],
['Letizia Lamartire'],
['Ray Jiang'],
['nan'],
['nan'],
['Daniel Vernon'],
['Steve Gukas'],
['Tim Johnson'],
['nan'],
['nan'],
['nan'],
['Vishwesh Krishnamoorthy'],
['nan'],
['nan'],
['Zack Snyder'],
['nan'],
['nan'],
['Uduak-Obong Patrick'],
['nan'],
['Mohamed Diab'],
['Amr Salama'],
['Christopher Amos'],
['Prakash Satam'],
['nan'],
['Robert Rodriguez'],
['Milan Luthria'],
['David Ayer'],
['Eshom Nelms', 'Ian Nelms'],
['James Moll'],
['nan'],
['nan'],
['James Redford'],
['Kaashvie Nair'],
['J.D. Dillard'],
['Nikhil Pherwani'],
['nan'],
['Mae Czarina Cruz'],
['nan'],
['Praveen Kandregula'],
['Cecilia Verheyden'],
['Daniel Minahan'],
['nan'],
['Donovan Marsh'],
['Brent Dawes'],
['nan'],
```

['nan'],
['Leli Maki'],
['Uzodinma Okpechi'],
['Daniel Prochaska'],
['nan'],
['Joe Wright'],
['nan'],
['Matthew Vaughn'],
['Aditya Kripalani'],
['Adriano Rudiman'],
['David Pablos'],
['Alexandre Aja'],
['nan'],
['Cai Cong'],
['Samuel Olatunji'],
['Ramon Térmens'],
['nan'],
['Svetlana Cvetko'],
['nan'],
['Martin Prakkat'],
['Baran bo Odar'],
['Zhang Chong'],
['Yılmaz Erdoğan'],
['Shantrelle P. Lewis'],
['nan'],
['Ivan Ayr'],
['Anthony Mandler'],
['Vijay Roche'],
['nan'],
['Stanley Menino D'Costa'],
['Jennifer Brea'],
['Julia von Heinz'],
['Niels Arden Oplev'],
['Don Argott', 'Sheena M. Joyce'],
['nan'],
['Joshua Zeman'],
['nan'],
['nan'],
['Duncan Skiles'],
['nan'],
['Sean McNamara'],
['nan'],
['Vondie Curtis-Hall'],
['nan'],
['Robert Radler'],
['Roel Reiné'],
['Todd Phillips'],
['Dean Parisot'],
['Paul Greengrass'],
['Lasse Hallström'],
['Justin Kelly'],
['Eric Darnell', 'Tom McGrath', 'Conrad Vernon'],
['nan'],
['Suhas Kadav'],
['Suhas Kadav'],
['Suhas Kadav'],
['Suhas Kadav'],
['Suhas Kadav'],
['Clint Eastwood'],
['Jeff Wadlow'],

```

['Charles Martin'],
['Stella Corradi'],
['Roland Emmerich'],
['Kevin Macdonald'],
['Ann Deborah Fishman'],
['Chris Gorak'],
['Peter Jackson'],
['Roger Kumble'],
['Jonathan Lynn'],
['Courtney Hunt'],
['Pierre Greco', 'Nancy Florence Savard'],
['Andrew Davis'],
['Kevin Smith'],
['nan'],
['Tosin Igho'],
['Chaitanya Tamhane'],
['Oriol Paulo'],
['Mike Rianda', 'Jeff Rowe'],
['Johannes Roberts'],
['nan'],
['Robert Pulcini', 'Shari Springer Berman'],
['nan'],
['Pedro Antonio'],
['nan'],
['nan'],
['John Wells'],
['Jonathan Liebesman'],
['Maria Pulera'],
['nan'],
['Santhosh Viswanath'],
['Seema Pahwa'],
['nan'],
['Ozan Açıktan'],
['Meltem Bozoflu'],
['Hakan Algül'],
['Selçuk Aydemir', 'Birkan Pusa'],
['Selçuk Aydemir'],
['Ömer Faruk Sorak'],
['Şenol Sönmez'],
['Alexis Morante'],
['Burak Aksak'],
['Kıvanç Baruönü'],
['Kıvanç Baruönü'],
['Rindala Kodeih'],
['Kongkiat Khomsiri'],
['Bedran Güzel'],
['Hakan Algül'],
['Marwan Nabil'],
['MIKIKO', 'Daito Manabe'],
['nan'],
['Kayode Kasum'],
['Yılmaz Erdoğan', 'Ömer Faruk Sorak'],
['Takashi Shimizu'],
['nan'],
['nan'],
['Joe Penna'],
...]
```

```
In [74]: df_new= pd.DataFrame(constraint, index=data['title'])
df_new
```

Out [74]:

	0	1	2	3	4	5	6	7	8	9	
title											
Dick Johnson Is Dead	Kirsten Johnson	None	None	None	None	None	None	None	None	None	N
Blood & Water	nan	None	None	None	None	None	None	None	None	None	N
Ganglands	Julien Leclercq	None	None	None	None	None	None	None	None	None	N
Jailbirds New Orleans	nan	None	None	None	None	None	None	None	None	None	N
Kota Factory	nan	None	None	None	None	None	None	None	None	None	N
...
Zodiac	David Fincher	None	None	None	None	None	None	None	None	None	N
Zombie Dumb	nan	None	None	None	None	None	None	None	None	None	N
Zombieland	Ruben Fleischer	None	None	None	None	None	None	None	None	None	N
Zoom	Peter Hewitt	None	None	None	None	None	None	None	None	None	N
Zubaan	Mozez Singh	None	None	None	None	None	None	None	None	None	N

8807 rows x 13 columns

In [84]: `df_director= df_new.stack().to_frame().reset_index()`In [85]: `print(df_director)`

```

      title  level_1
0  Dick Johnson Is Dead      0  Kirsten Johnson
1      Blood & Water      0             nan
2      Ganglands      0  Julien Leclercq
3  Jailbirds New Orleans      0             nan
4      Kota Factory      0             nan
...      ...      ...
9607      Zodiac      0  David Fincher
9608      Zombie Dumb      0             nan
9609      Zombieland      0  Ruben Fleischer
9610      Zoom      0  Peter Hewitt
9611      Zubaan      0  Mozez Singh

```

[9612 rows x 3 columns]

In [92]: `df_director= pd.DataFrame({'title': df_director['title'], 'director': df_director`

Out [92]:

	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	Mozes Singh

9612 rows × 2 columns

```
In [96]: constraint= data['country'].apply(lambda x: str(x).split(', ').to_list())
df_country= pd.DataFrame(constraint, index=data['title'])
df_country=df_country.stack().to_frame().reset_index()
df_country= df_country[['title', 0]]
df_country= df_country.rename(columns={0: 'country'})
df_country
```

Out [96]:

	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 [97]: constraint= data['listed_in'].apply(lambda x: str(x).split(', ').to_list())
df_listed_in= pd.DataFrame(constraint, index=data['title'])
df_listed_in=df_listed_in.stack().to_frame().reset_index()
df_listed_in= df_listed_in[['title', 0]]
```

```
df_listed_in= df_listed_in.rename(columns={0:'listed_in'})
df_listed_in
```

Out [97]:

	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 [102... new_data= df_cast.merge(df_director, on='title').merge(df_country, on='ti
```

```
In [108... data_new= new_data.merge(data[['title', 'show_id', 'type', 'date_added', '
```

```
In [112... data.columns
```

```
Out[112... Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_
added',
      'release_year', 'rating', 'duration', 'listed_in', 'descriptio
n'],
      dtype='object')
```

```
In [ ]: # Getting Final data after unpacking all the csv series and merging them
```

```
In [113... final_data= data_new[['show_id', 'type', 'title', 'director', 'cast', 'co
      'release_year', 'rating', 'duration', 'listed_in', 'description']]
final_data
```

Out [113...

	show_id	type	title	director	cast	country	date_added	rele
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	nan	United States	September 25, 2021	
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	
2	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	
3	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	
4	s2	TV Show	Blood & Water	nan	Khosi Ngema	South Africa	September 24, 2021	
...
201986	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	
201987	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	
201988	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	
201989	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	
201990	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	

show_id	type	title	director	cast	country	date_added	rele
---------	------	-------	----------	------	---------	------------	------

201991 rows × 12 columns

In [117... `final_data.info()`]

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 201991 entries, 0 to 201990
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   show_id                201991 non-null object
1   type                  201991 non-null object
2   title                 201991 non-null object
3   director              201991 non-null object
4   cast                  201991 non-null object
5   country               201991 non-null object
6   date_added            201833 non-null object
7   release_year          201991 non-null int64
8   rating                201924 non-null object
9   duration              201988 non-null object
10  listed_in             201991 non-null object
11  description            201991 non-null object
dtypes: int64(1), object(11)
memory usage: 18.5+ MB
```

In []: `# Changing the data type of 'date_added' from 'object' to 'date_time'`

In [121... `final_data['date_added'] = pd.to_datetime(final_data['date_added'], errors`

In [122... `final_data.info()`]

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 201991 entries, 0 to 201990
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   show_id                201991 non-null object
1   type                  201991 non-null object
2   title                 201991 non-null object
3   director              201991 non-null object
4   cast                  201991 non-null object
5   country               201991 non-null object
6   date_added            200245 non-null datetime64[ns]
7   release_year          201991 non-null int64
8   rating                201924 non-null object
9   duration              201988 non-null object
10  listed_in             201991 non-null object
11  description            201991 non-null object
dtypes: datetime64[ns](1), int64(1), object(10)
memory usage: 18.5+ MB
```

In []: `# Value_counts & Unique attributes`

In []: `# Comparison of tv shows vs movies`

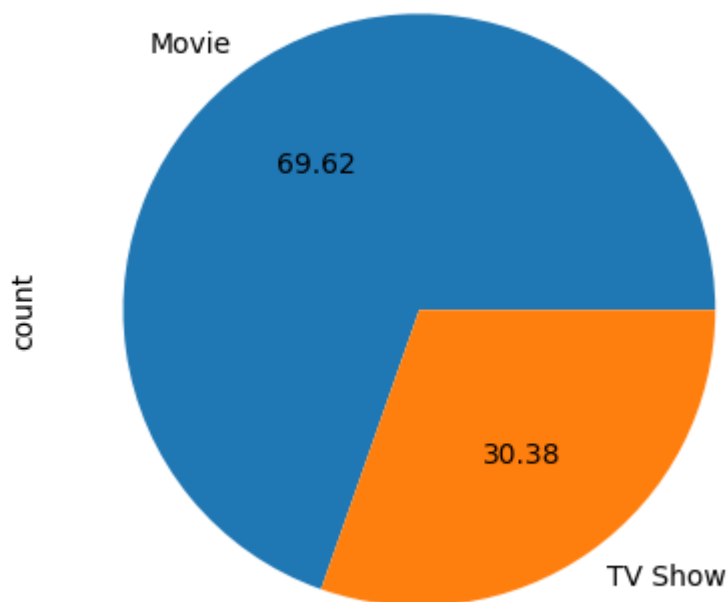
```
In [124... data['type'].value_counts()
```

```
Out[124... type
Movie      6131
TV Show    2676
Name: count, dtype: int64
```

```
In [125... import seaborn as sns
```

```
In [128... data['type'].value_counts().plot(kind='pie', autopct='%.2f')
```

```
Out[128... <Axes: ylabel='count'>
```



```
In [ ]: # variation of content type over years
```

```
In [152... final_data['year_added'] = final_data['date_added'].dt.year
```

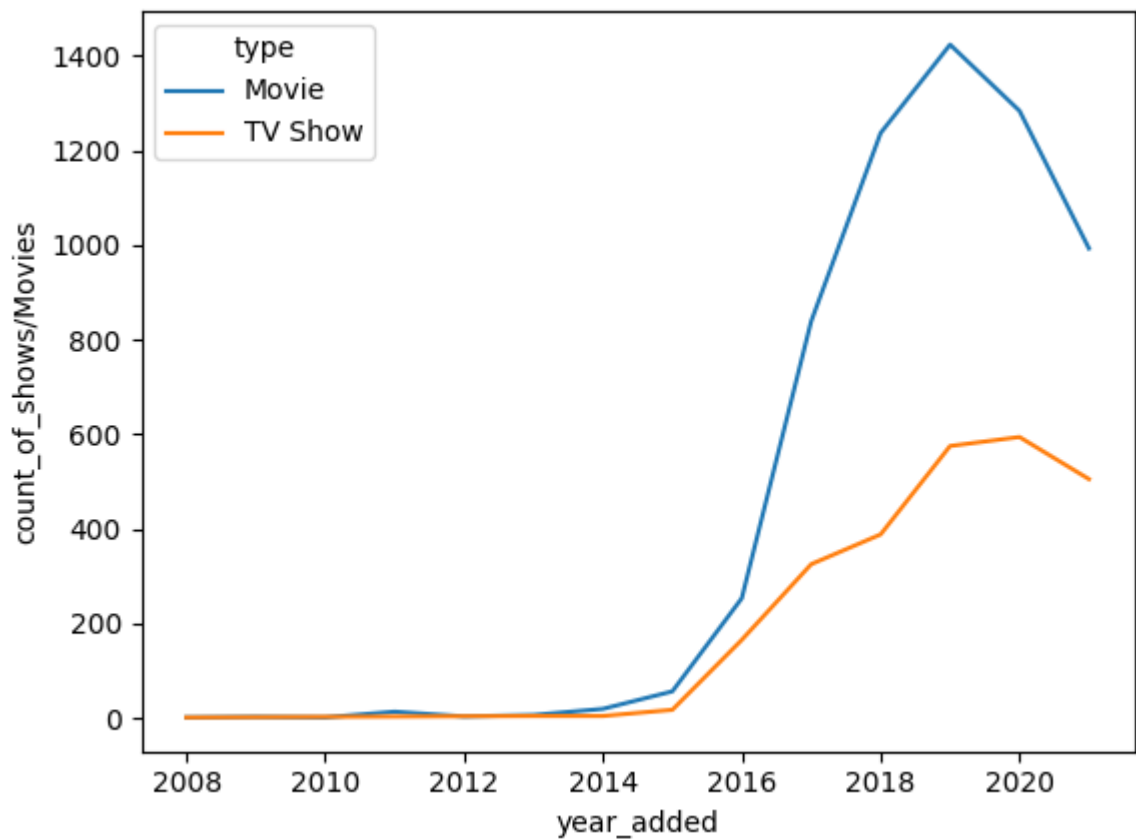
```
In [155... final_data['year_added']
```

```
Out[155... 0      2021.0
1      2021.0
2      2021.0
3      2021.0
4      2021.0
...
201986  2019.0
201987  2019.0
201988  2019.0
201989  2019.0
201990  2019.0
Name: year_added, Length: 201991, dtype: float64
```

```
In [162... yearly_count = final_data.groupby(['year_added', 'type'])['show_id'].nunique
yearly_count.rename(columns={'show_id': 'count_of_shows/Movies'}, inplace=T
```

```
In [163... sns.lineplot(data=yearly_count, x='year_added', y='count_of_shows/Movies')
```

```
Out[163... <Axes: xlabel='year_added', ylabel='count_of_shows/Movies'>
```



```
In [ ]: # Movies release per year
```

```
In [192... Movies= final_data.loc[final_data['type']=='Movie', ['show_id','release_y  
Movies
```

```
Out[192... show_id release_year
```

0	s1	2020
159	s7	2021
160	s7	2021
161	s7	2021
162	s7	2021
...
201986	s8807	2015
201987	s8807	2015
201988	s8807	2015
201989	s8807	2015
201990	s8807	2015

145843 rows x 2 columns

```
In [193... Movies= Movies.groupby('release_year')['show_id'].nunique().reset_index()

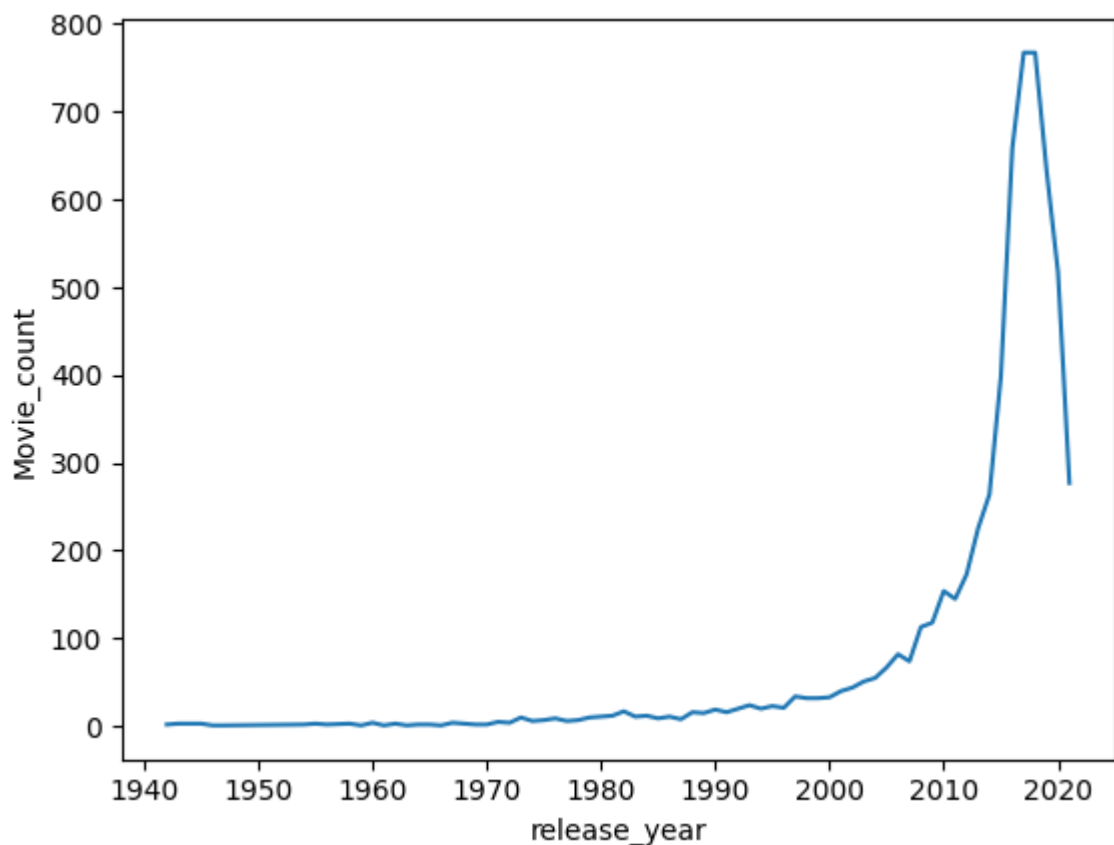
Movies.rename(columns={'show_id': 'Movie_count'}, inplace=True)
Movies
```

```
Out [193...      release_year  Movie_count
0          1942             2
1          1943             3
2          1944             3
3          1945             3
4          1946             1
...           ...           ...
68         2017          767
69         2018          767
70         2019          633
71         2020          517
72         2021          277
```

73 rows × 2 columns

```
In [194... sns.lineplot(data=Movies, x='release_year', y='Movie_count' )
```

```
Out [194... <Axes: xlabel='release_year', ylabel='Movie_count'>
```



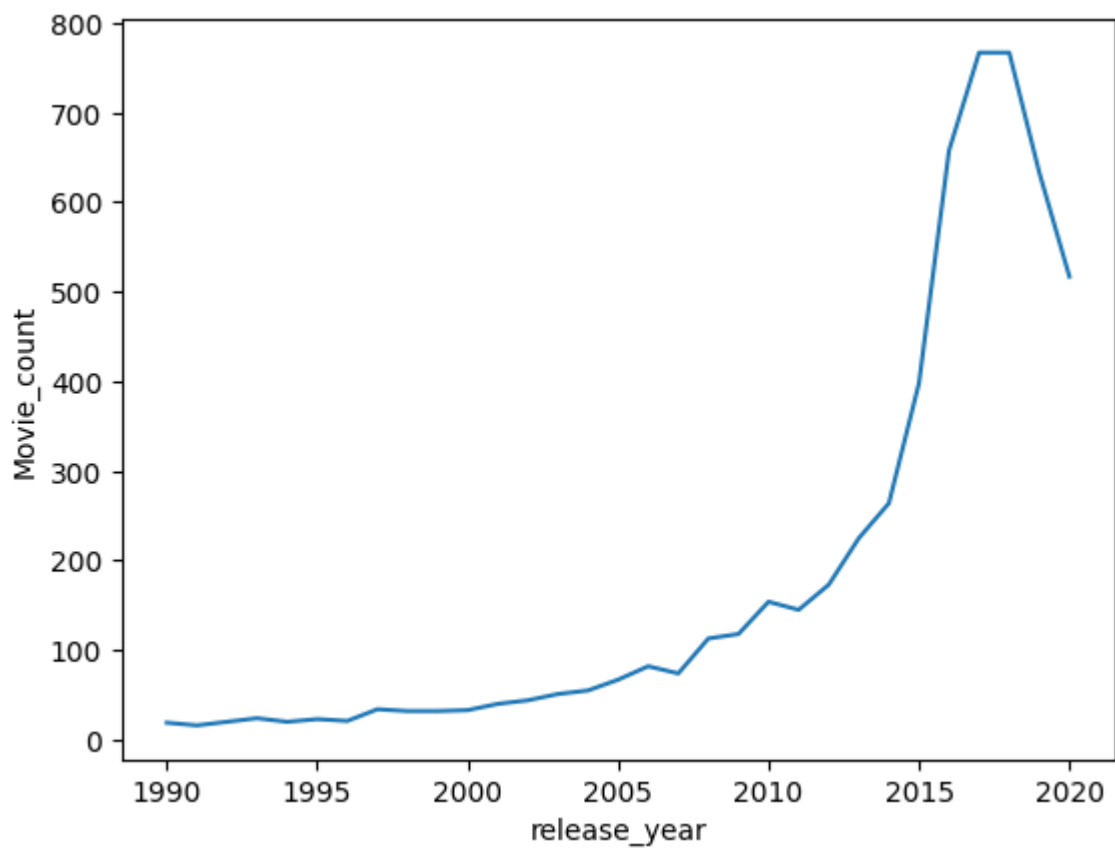
```
In [203... last_thirty_year_releases= Movies.loc[Movies['release_year'].between(1990, 2020)]
last_thirty_year_releases
```

```
Out[203... 
```

	release_year	Movie_count
41	1990	19
42	1991	16
43	1992	20
44	1993	24
45	1994	20
46	1995	23
47	1996	21
48	1997	34
49	1998	32
50	1999	32
51	2000	33
52	2001	40
53	2002	44
54	2003	51
55	2004	55
56	2005	67
57	2006	82
58	2007	74
59	2008	113
60	2009	118
61	2010	154
62	2011	145
63	2012	173
64	2013	225
65	2014	264
66	2015	398
67	2016	658
68	2017	767
69	2018	767
70	2019	633
71	2020	517

```
In [204... sns.lineplot(data=last_thirty_year_releases, x='release_year', y='Movie_c
```

```
Out[204... <Axes: xlabel='release_year', ylabel='Movie_count'>
```



```
In [ ]: # Content available in different countries
```

```
In [205... final_data.head()
```

Out [205...

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	nan	United States	2021-09-25	2020
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	2021-09-24	2021
2	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	2021-09-24	2021
3	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	2021-09-24	2021
4	s2	TV Show	Blood & Water	nan	Khosi Ngema	South Africa	2021-09-24	2021

```
In [222...] country_wise_content= final_data.groupby(['country','listed_in'])['title'
```

```
In [223...] country_wise_content.rename(columns={'title':'count'},inplace=True)
```

```
In [241...] country_wise_content.dropna(axis=1,how='any')
```

Out [241...

	country	listed_in	count
0		Dramas	1
1		Independent Movies	1
2		International Movies	1
3		International TV Shows	1
4		TV Dramas	1
...
1459	nan	TV Sci-Fi & Fantasy	5
1460	nan	TV Shows	7
1461	nan	TV Thrillers	4
1462	nan	Teen TV Shows	3
1463	nan	Thrillers	28

1464 rows x 3 columns

In [244...

```
country_wise_content
```

Out [244...

	country	listed_in	count
0		Dramas	1
1		Independent Movies	1
2		International Movies	1
3		International TV Shows	1
4		TV Dramas	1
...
1459	nan	TV Sci-Fi & Fantasy	5
1460	nan	TV Shows	7
1461	nan	TV Thrillers	4
1462	nan	Teen TV Shows	3
1463	nan	Thrillers	28

1464 rows x 3 columns

In []:

```
# Best month for tv-shows
```

In [263...

```
data['month']= data['date_added'].dt.month
```

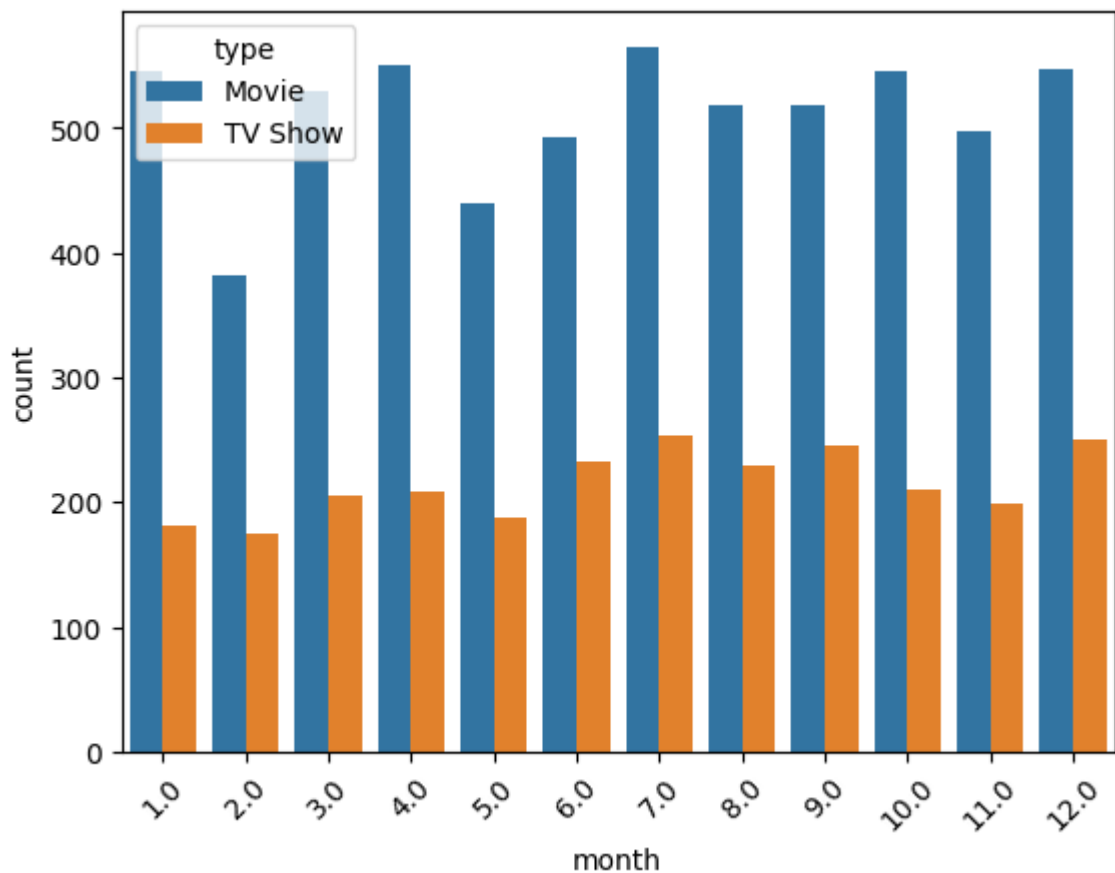
In [264...

```
data['month']
```



```
Out [264...] 0      9.0
              1      9.0
              2      9.0
              3      9.0
              4      9.0
              ...
            8802    11.0
            8803     7.0
            8804    11.0
            8805     1.0
            8806     3.0
            Name: month, Length: 8807, dtype: float64
```

```
In [265...] sns.countplot(x='month', hue='type', data=data)
            plt.xticks(rotation=45)
            plt.show()
```



```
In [ ]: # Most Popular Director and Actor
```

```
In [281...] popular_director= final_data.groupby(['director'])['title'].nunique().res
            popular_director.sort_values('release_count', ascending=False)
```

Out [281]...

	director	release_count
4978	nan	2634
3749	Rajiv Chilaka	22
1906	Jan Suter	21
3800	Raúl Campos	19
4457	Suhas Kadav	16
...
635	Brandon Camp	1
2295	Juan Antin	1
2296	Juan Antonio de la Riva	1
2297	Juan Camilo Pinzon	1
2957	María Jose Cuevas	1

4994 rows × 2 columns

In [282]...

```
popular_actor= final_data.groupby(['cast'])['title'].nunique().reset_index()
popular_actor.sort_values('show_count', ascending=False)
```

Out [282]...

	cast	show_count
36316	nan	825
2833	Anupam Kher	43
30489	Shah Rukh Khan	35
16697	Julie Tejjwani	33
24215	Naseeruddin Shah	32
...
14221	Jamie Lee	1
14219	Jamie Kenna	1
14218	Jamie Kaler	1
14217	Jamie Johnston	1
36439	Şöpe Dirisù	1

36440 rows × 2 columns

```
In [ ]: # Most Popular Director and Actor country-wise
```

In [288]...

```
director_country_wise = final_data.groupby(['country','director'])['title']
```

In [297]...

```
director_country_wise=director_country_wise.dropna(subset=['country'])
```

In [298]...

```
director_country_wise= director_country_wise.reset_index(drop=True)
```

```
In [302... director_country_wise.sort_values('release_count', ascending=False)
```

```
Out[302...

```

	country	director	release_count
6682	United States	nan	940
7094	nan	nan	409
4376	United Kingdom	nan	267
2733	Japan	nan	188
3433	South Korea	nan	161
...
2524	Italy	Andrea Molaioli	1
2523	Italy	Alice Rohrwacher	1
2522	Italy	Alex Infascelli	1
2521	Italy	Alex Holmes	1
3547	Spain	Manolo Caro	1

7095 rows x 3 columns

```
In [ ]: # Rating-wise most produced content
```

```
In [305... rating_wise_release= final_data.groupby('listed_in')['title'].nunique().r
```

```
In [306... rating_wise_release.sort_values('release_count', ascending=False)
```

Out [306...

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

	listed_in	release_count
9	Cult Movies	71
40	Teen TV Shows	69
13	Faith & Spirituality	65
39	TV Thrillers	57
21	Movies	57
31	Stand-Up Comedy & Talk Shows	56
5	Classic & Cult TV	28
38	TV Shows	16

```
In [ ]: # average duration for movies and TV shows
```

```
In [355... # data['duration']= data['duration'].apply(lambda x: int(str(x).split(' '
nanDuration = data[ (data['duration'] == 'nan')].index
data.drop(nanDuration , inplace=True)
# data['duration'].info()
data['duration']= data['duration'].apply(lambda x: int(str(x).split(' ')[
data['duration']
```

```
Out[355... 0      90
1       2
2       1
3       1
4       2
...
8802    158
8803     2
8804     88
8805     88
8806    111
Name: duration, Length: 8804, dtype: int64
```

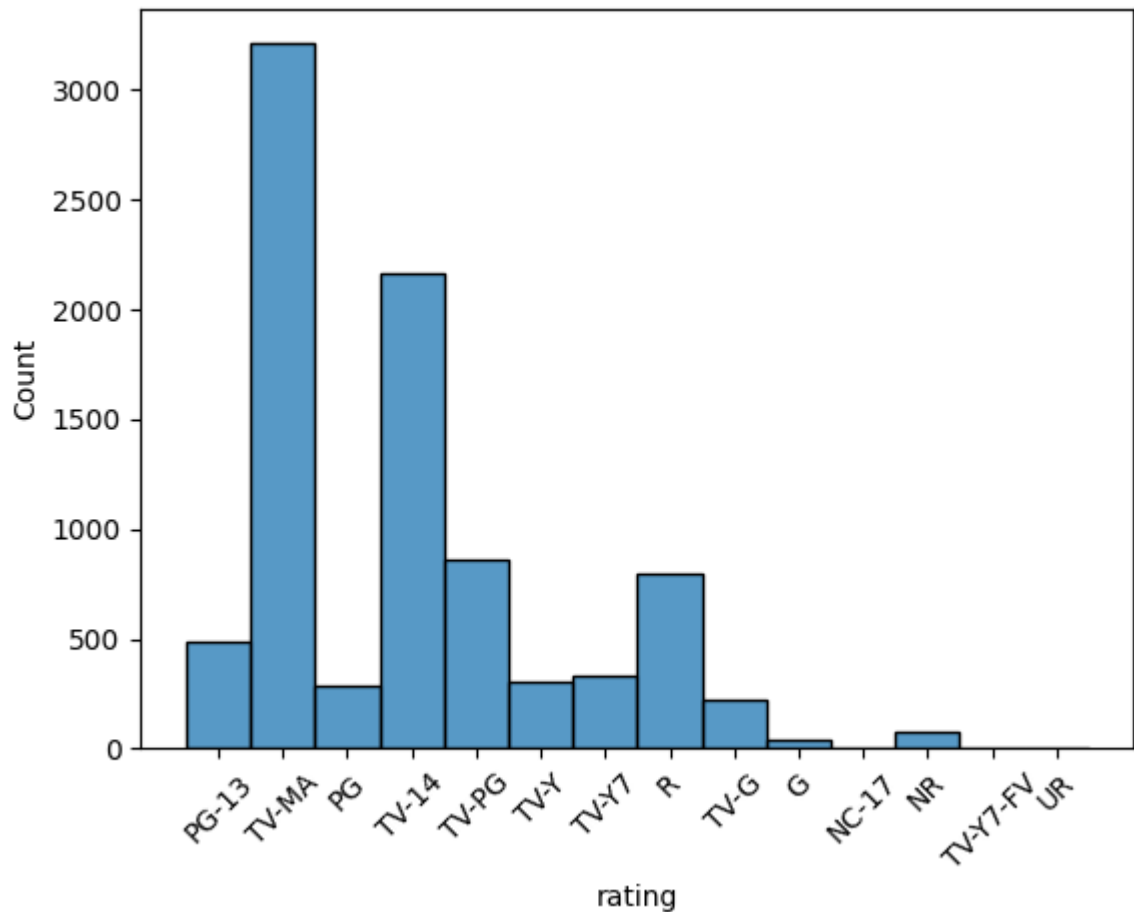
```
In [356... mean_duration= data.groupby('type')['duration'].mean().reset_index()
```

```
In [357... mean_duration
```

```
Out[357...      type  duration
0  Movie  99.577187
1  TV Show  1.764948
```

```
In [ ]: # Univariate analysis- rating histogram
```

```
In [360... sns.histplot(data=data, x='rating')
plt.xticks(rotation=45)
plt.show()
```



In []: `# Bivariate Analysis`

In [364...]: `Movies`

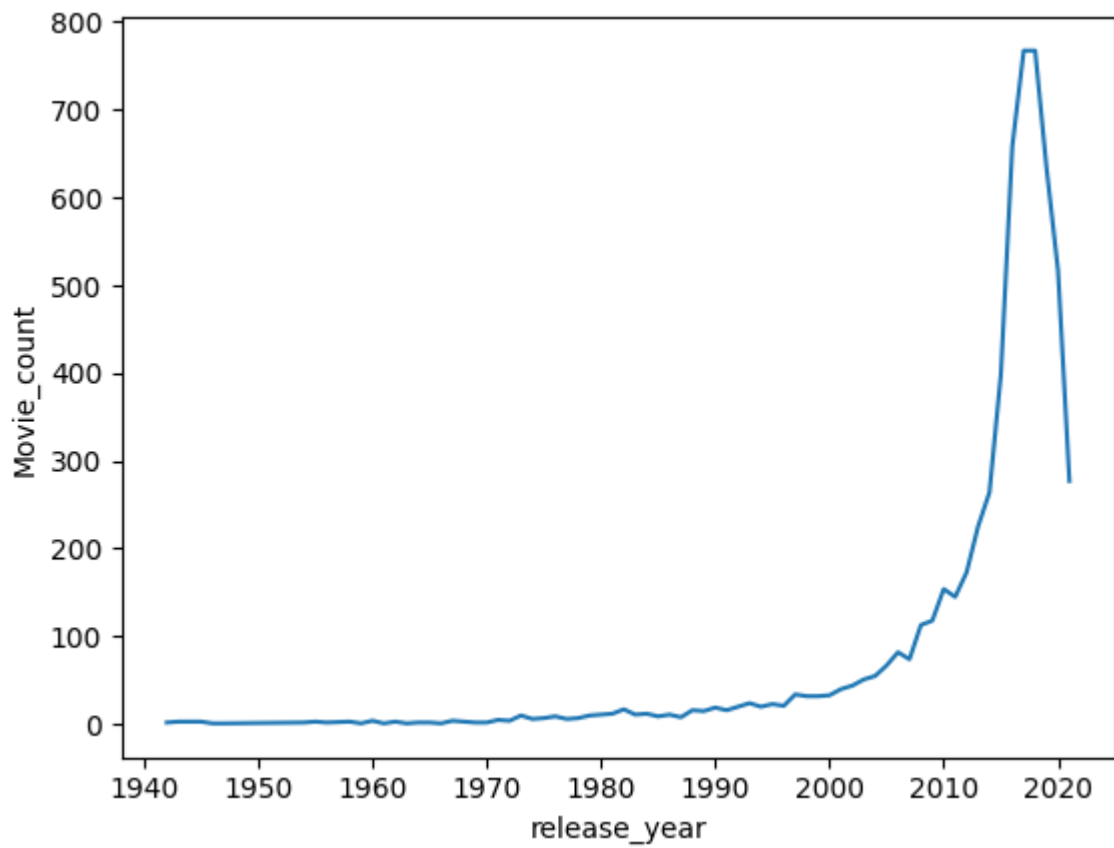
Out [364...]:

	release_year	Movie_count
0	1942	2
1	1943	3
2	1944	3
3	1945	3
4	1946	1
...
68	2017	767
69	2018	767
70	2019	633
71	2020	517
72	2021	277

73 rows x 2 columns

In [363...]: `sns.lineplot(data=Movies, x='release_year', y='Movie_count')`

Out[363... <Axes: xlabel='release_year', ylabel='Movie_count'>



```
In [ ]: # Multivariate Analysis
```

```
In [369... Anupam_Kher_data=final_data.loc[final_data['cast']=='Anupam Kher']
```

```
In [370... Anupam_Kher_data
```

Out [370...

	show_id	type	title	director	cast	country	date_added	release_date
4722	s193	Movie	C Kkompany	Sachin Yardi	Anupam Kher	India	2021-08-27	
4723	s193	Movie	C Kkompany	Sachin Yardi	Anupam Kher	India	2021-08-27	
4724	s193	Movie	C Kkompany	Sachin Yardi	Anupam Kher	India	2021-08-27	
5027	s203	Movie	Kyaa Kool Hai Hum	Sangeeth Sivan	Anupam Kher	India	2021-08-27	
5028	s203	Movie	Kyaa Kool Hai Hum	Sangeeth Sivan	Anupam Kher	India	2021-08-27	
...
201004	s8770	Movie	Y.M.I.: Yeh Mera India	N. Chandra	Anupam Kher	India	2018-03-01	
201005	s8770	Movie	Y.M.I.: Yeh Mera India	N. Chandra	Anupam Kher	India	2018-03-01	
201132	s8773	Movie	Yamla Pagla Deewana 2	Sangeeth Sivan	Anupam Kher	India	2017-05-01	
201133	s8773	Movie	Yamla Pagla Deewana 2	Sangeeth Sivan	Anupam Kher	India	2017-05-01	
201134	s8773	Movie	Yamla Pagla Deewana 2	Sangeeth Sivan	Anupam Kher	India	2017-05-01	

127 rows × 14 columns

