

Unzip the file

```
In [2]: import zipfile
import os

# Specify the path to the uploaded zip file on your desktop
zip_file_path = os.path.join(os.path.expanduser("~"), "Desktop", "sample_project", "Netflix_EDA", "Netflix", "Netflix_Data.zip")

# Specify the destination directory for extraction
extracted_dir_path = os.path.join(os.path.expanduser("~"), "Desktop", "sample_project", "Netflix_EDA", "Netflix")

# Create the destination directory if it doesn't exist
os.makedirs(extracted_dir_path, exist_ok=True)

# Unzip the file
with zipfile.ZipFile(zip_file_path, 'r') as zip_ref:
    zip_ref.extractall(extracted_dir_path)

print(f"File '{zip_file_path}' successfully extracted to '{extracted_dir_path}'.")
```

File 'C:\Users\RDS Kolkata\Desktop\sample_project\Netflix_EDA\Netflix\Netflix_Data.zip' successfully extracted to 'C:\Users\RDS Kolkata\Desktop\sample_project\Netflix_EDA\Netflix'.

```
In [3]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Reading And Inspecting Data

```
In [4]: df=pd.read_csv("Netflix_EDA/Netflix/netflix_titles_2021.csv")
```

```
In [5]: df.head()
```

```
Out[5]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mablane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Goloas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

```
In [6]: 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 [7]: df.describe()

Out[7]:
```

	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

The dataset consist of following columns

- show_id: Gives the information about show id.
- type: Gives information about 2 different unique values one is TV Show and another is Movie.
- title: Gives information about the title of Movie or TV Show.
- director: Gives information about the director who directed the Movie or TV Show.
- cast: Gives information about the cast who plays role in Movie or TV Show.
- country: Gives information about the Name of country.
- date_added: Gives information about the tv shows or movie released.
- release_year: Gives information about the year when Movie or TV Show was released.
- rating: Gives information about the Movie or TV Show are in which category (eg like the movies are only for students, or adults, etc).
- duration: Gives information about the duration of Movie or TV Show.
- listed_in: Gives information about the genre of Movie or TV Show.
- description: Gives information about the description of Movie or TV Show.

```
In [8]: print(f'Dataset has {df.shape[0]} rows and {df.shape[1]} columns')

Dataset has 8807 rows and 12 columns
```

```
In [9]: df.head()

Out[9]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Goloas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

```
In [10]: zip_file_path

Out[10]: 'C:\\Users\\RDS Kolkata\\Desktop\\sample_project\\Netflix_EDA\\Netflix\\Netflix_Data.zip'

In [11]: df['date_added'].head(20)

Out[11]:
```

0	September 25, 2021
1	September 24, 2021
2	September 24, 2021
3	September 24, 2021
4	September 24, 2021
5	September 24, 2021
6	September 24, 2021
7	September 24, 2021
8	September 24, 2021
9	September 24, 2021
10	September 24, 2021
11	September 23, 2021
12	September 23, 2021
13	September 22, 2021
14	September 22, 2021
15	September 22, 2021
16	September 22, 2021
17	September 22, 2021
18	September 22, 2021
19	September 22, 2021

Name: date_added, dtype: object

```
In [12]: df['rating']

Out[12]: 0      PG-13
1      TV-MA
2      TV-MA
3      TV-MA
4      TV-MA
...
8802      R
8803      TV-Y7
8804      R
8805      PG
8806      TV-14
Name: rating, Length: 8807, dtype: object
```

Data Cleaning

Handling Missing Values

```
In [13]: # % of missing values
df.isnull().mean()*100

Out[13]: show_id      0.000000
type      0.000000
title      0.000000
director   29.908028
cast       9.367549
country    9.435676
date_added 0.113546
release_year 0.000000
rating     0.045418
duration   0.034064
listed_in  0.000000
description 0.000000
dtype: float64
```

```
In [14]: df[df['director'].isnull()]

Out[14]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...
10	s11	TV Show	Vendetta: Truth, Lies and The Mafia	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, Docuseries, International TV S...	Sicily boasts a bold "Anti-Mafia" coalition. B...
14	s15	TV Show	Crime Stories: India Detectives	NaN	NaN	NaN	September 22, 2021	2021	TV-MA	1 Season	British TV Shows, Crime TV Shows, Docuseries	Cameras following Bengaluru police on the job ...
...
8795	s8796	TV Show	Yu-Gi-Oh! Arc-V	NaN	Mike Liscio, Emily Bauer, Billy Bob Thompson, ...	Japan, Canada	May 1, 2018	2015	TV-Y7	2 Seasons	Anime Series, Kids' TV	Now that he's discovered the Pendulum Summonin...
8796	s8797	TV Show	Yunus Emre	NaN	Gökhan Atalay, Payidar Tüfekçioğlu, Baran Akbu...	Turkey	January 17, 2017	2016	TV-PG	2 Seasons	International TV Shows, TV Dramas	During the Mongol invasions, Yunus Emre leaves...
8797	s8798	TV Show	Zak Storm	NaN	Michael Johnston, Jessica Gee-George, Christin...	United States, France, South Korea, Indonesia	September 13, 2018	2016	TV-Y7	3 Seasons	Kids' TV	Teen surfer Zak Storm is mysteriously transpor...
8800	s8801	TV Show	Zindagi Gulzar Hai	NaN	Sanam Saeed, Fawad Khan, Ayesha Omer, Mehreen ...	Pakistan	December 15, 2016	2012	TV-PG	1 Season	International TV Shows, Romantic TV Shows, TV ...	Strong-willed, middle-class Kashaf and carefre...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...

2634 rows × 12 columns

```
In [15]: df['director']=df['director'].fillna('unknown')
df['director'].isnull().sum()

Out[15]: 0
```

In [16]:

df[df['cast'].isnull()]

Out[16]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
3	s4	TV Show	Jailbirds New Orleans	unknown	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
10	s11	TV Show	Vendetta: Truth, Lies and The Mafia	unknown	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, Docuseries, International TV S...	Sicily boasts a bold "Anti-Mafia" coalition. B...
14	s15	TV Show	Crime Stories: India Detectives	unknown	NaN	NaN	September 22, 2021	2021	TV-MA	1 Season	British TV Shows, Crime TV Shows, Docuseries	Cameras following Bengaluru police on the job ...
16	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	Pedro de Echave García, Pablo Azorin Williams	NaN	NaN	September 22, 2021	2020	TV-MA	67 min	Documentaries, International Movies	Declassified documents reveal the post-WWII li...
...
8746	s8747	Movie	Winnie	Pascale Lamche	NaN	France, Netherlands, South Africa, Finland	February 26, 2018	2017	TV-14	85 min	Documentaries, International Movies	Winnie Mandela speaks about her extraordinary ...
8755	s8756	TV Show	Women Behind Bars	unknown	NaN	United States	November 1, 2016	2010	TV-14	3 Seasons	Crime TV Shows, Docuseries	This reality series recounts true stories of w...
8756	s8757	Movie	Woodstock	Barak Goodman	NaN	United States	August 13, 2019	2019	TV-MA	97 min	Documentaries, Music & Musicals	For the 50th anniversary of the legendary Wood...
8763	s8764	Movie	WWII: Report from the Aleutians	John Huston	NaN	United States	March 31, 2017	1943	TV-PG	45 min	Documentaries	Filmmaker John Huston narrates this Oscar-nomi...
8803	s8804	TV Show	Zombie Dumb	unknown	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...

825 rows × 12 columns

In [17]:

df['cast'] = df['cast'].fillna('unknown')
df['cast'].isnull().sum()

Out[17]:

0

In [18]:

df[df['country'].isnull()]

Out[18]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	unknown	unknown	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	The arrival of a charismatic young priest brin...
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	91 min	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...
10	s11	TV Show	Vendetta: Truth, Lies and The Mafia	unknown	unknown	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, Docuseries, International TV S...	Sicily boasts a bold "Anti-Mafia" coalition. B...
...
8718	s8719	Movie	Westside vs. the World	Michael Fahey	Ron Perlman, Louie Simmons	NaN	August 9, 2019	2019	TV-MA	96 min	Documentaries, Sports Movies	A look into the journey of influential strengt...
8759	s8760	Movie	World's Weirdest Homes	Storm Theunissen	Charlie Luxton	NaN	February 1, 2019	2015	TV-PG	49 min	Movies	From a bubble-shaped palace to an island built...
8783	s8784	TV Show	Yoko	unknown	Eileen Stevens, Alyson Leigh Rosenfeld, Sarah ...	NaN	June 23, 2018	2016	TV-Y	1 Season	Kids' TV	Friends Mai, Oto and Vik's games at the park b...
8785	s8786	TV Show	YOM	unknown	Sairaj, Devyani Dagaonkar, Ketan Singh, Mayur ...	NaN	June 7, 2018	2016	TV-Y7	1 Season	Kids' TV	With the mind of a human being, and the body o...
8803	s8804	TV Show	Zombie Dumb	unknown	unknown	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...

831 rows × 12 columns

In [19]:

df['country'] = df['country'].fillna('unknown')
df['country'].isnull().sum()

Out[19]:

0

In [20]:

df[df['country'].isnull()]

Out[20]:

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

```
In [21]: df[df['date_added'].isnull()]

Out[21]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
6066	s6067	TV Show	A Young Doctor's Notebook and Other Stories	unknown	Daniel Radcliffe, Jon Hamm, Adam Godley, Chris...	United Kingdom	NaN	2013	TV-MA	2 Seasons	British TV Shows, TV Comedies, TV Dramas	Set during the Russian Revolution, this comic ...
6174	s6175	TV Show	Anthony Bourdain: Parts Unknown	unknown	Anthony Bourdain	United States	NaN	2018	TV-PG	5 Seasons	Docuseries	This CNN original series has chef Anthony Bour...
6795	s6796	TV Show	Frasier	unknown	Kelsey Grammer, Jane Leeves, David Hyde Pierce...	United States	NaN	2003	TV-PG	11 Seasons	Classic & Cult TV, TV Comedies	Frasier Crane is a snooty but lovable Seattle ...
6806	s6807	TV Show	Friends	unknown	Jennifer Aniston, Courteney Cox, Lisa Kudrow, ...	United States	NaN	2003	TV-14	10 Seasons	Classic & Cult TV, TV Comedies	This hit sitcom follows the merry misadventure...
6901	s6902	TV Show	Gunslinger Girl	unknown	Yuuka Nanri, Kanako Mitsuhashi, Eri Sendai, Am...	Japan	NaN	2008	TV-14	2 Seasons	Anime Series, Crime TV Shows	On the surface, the Social Welfare Agency appe...
7196	s7197	TV Show	Kikoriki	unknown	Igor Dmitriev	unknown	NaN	2010	TV-Y	2 Seasons	Kids' TV	A wacky rabbit and his gang of animal pals hav...
7254	s7255	TV Show	La Familia P. Luche	unknown	Eugenio Derbez, Consuelo Duval, Luis Manuel Áv...	United States	NaN	2012	TV-14	3 Seasons	International TV Shows, Spanish-Language TV Sh...	This irreverent sitcom features Ludovico, Feder...
7406	s7407	TV Show	Maron	unknown	Marc Maron, Judd Hirsch, Josh Brener, Nora Zeh...	United States	NaN	2016	TV-MA	4 Seasons	TV Comedies	Marc Maron stars as Marc Maron, who interviews...
7847	s7848	TV Show	Red vs. Blue	unknown	Burnie Burns, Jason Saldana, Gustavo Sorola, G...	United States	NaN	2015	NR	13 Seasons	TV Action & Adventure, TV Comedies, TV Sci-Fi ...	This parody of first-person shooter games, mil...
8182	s8183	TV Show	The Adventures of Figaro Pho	unknown	Luke Jurevicius, Craig Behenna, Charlotte Haml...	Australia	NaN	2015	TV-Y7	2 Seasons	Kids' TV, TV Comedies	Imagine your worst fears, then multiply them: ...

```
In [22]: df.dropna(inplace=True)
df.isnull().mean()*100

Out[22]:
```

show_id	0.0
type	0.0
title	0.0
director	0.0
cast	0.0
country	0.0
date_added	0.0
release_year	0.0
rating	0.0
duration	0.0
listed_in	0.0
description	0.0
dtype:	float64

Standarization

```
In [23]: df['country']=df['country'].apply(lambda x: x.capitalize())
df['country'].head()

Out[23]:
```

0	United states
1	South africa
2	Unknown
3	Unknown
4	India

Name: country, dtype: object

In [24]:

df.head()

Out[24]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	unknown	United states	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	unknown	Ama Qamata, Khosi Ngema, Gail Mablane, Thaban...	South africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Unknown	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	unknown	unknown	Unknown	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	unknown	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

In [25]:

df[['date_added','release_year']]

Out[25]:

	date_added	release_year
0	September 25, 2021	2020
1	September 24, 2021	2021
2	September 24, 2021	2021
3	September 24, 2021	2021
4	September 24, 2021	2021
...
8802	November 20, 2019	2007
8803	July 1, 2019	2018
8804	November 1, 2019	2009
8805	January 11, 2020	2006
8806	March 2, 2019	2015

8790 rows × 2 columns

In [26]:

Converting date_added into month & year
df['month_added'], df['year_added'] = pd.to_datetime(df['date_added']).dt.strftime('%B'), pd.to_datetime(df['date_added']).dt.

In [27]:

df[['month_added','year_added','release_year']]

Out[27]:

	month_added	year_added	release_year
0	September	2021	2020
1	September	2021	2021
2	September	2021	2021
3	September	2021	2021
4	September	2021	2021
...
8802	November	2019	2007
8803	July	2019	2018
8804	November	2019	2009
8805	January	2020	2006
8806	March	2019	2015

8790 rows × 3 columns

In [28]:

dropping column date_added
df.drop('date_added',axis=1,inplace=True)

In [29]:

df['year_added']=df['year_added'].astype(int)

```
In [30]: df[df['year_added']<df['release_year']]
```

Out[30]:

	show_id	type	title	director	cast	country	release_year	rating	duration	listed_in	description	month_added	year_added	
	1551	s1552	TV Show	Hilda	unknown	Bella Ramsey, Ameerah Falzon-Ojo, Oliver Nelso...	United kingdom, canada, united states	2021	TV-Y7	2 Seasons	Kids' TV	Fearless, free-spirited Hilda finds new friend...	December	2020
	1696	s1697	TV Show	Polly Pocket	unknown	Emily Tennant, Shannon Chan-Kent, Kazumi Evans...	Canada, united states, ireland	2021	TV-Y	2 Seasons	Kids' TV	After uncovering a magical locket that allows ...	November	2020
	2920	s2921	TV Show	Love Is Blind	unknown	Nick Lachey, Vanessa Lachey	United states	2021	TV-MA	1 Season	Reality TV, Romantic TV Shows	Nick and Vanessa Lachey host this social exper...	February	2020
	3168	s3169	TV Show	Fuller House	unknown	Candace Cameron Bure, Jodie Sweetin, Andrea Ba...	United states	2020	TV-PG	5 Seasons	TV Comedies	The Tanner family's adventures continue as DJ ...	December	2019
	3287	s3288	TV Show	Maradona in Mexico	unknown	Diego Armando Maradona	Argentina, united states, mexico	2020	TV-MA	1 Season	Docuseries, Spanish-Language TV Shows	In this docuseries, soccer great Diego Maradon...	November	2019
	3369	s3370	TV Show	BoJack Horseman	unknown	Will Arnett, Aaron Paul, Amy Sedaris, Alison B...	United states	2020	TV-MA	6 Seasons	TV Comedies	Meet the most beloved sitcom horse of the '90s...	October	2019
	3433	s3434	TV Show	The Hook Up Plan	unknown	Marc Ruchmann, Zita Hanrot, Sabrina Ouazani, J...	France	2020	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	When Parisian Elsa gets hung up on her ex, her...	October	2019
	4844	s4845	TV Show	Unbreakable Kimmy Schmidt	unknown	Ellie Kemper, Jane Krakowski, Tituss Burgess, ...	United states	2019	TV-14	4 Seasons	TV Comedies	When a woman is rescued from a doomsday cult a...	May	2018
	4845	s4846	TV Show	Arrested Development	unknown	Jason Bateman, Portia de Rossi, Will Arnett, M...	United states	2019	TV-MA	5 Seasons	TV Comedies	It's the Emmy-winning story of a wealthy famil...	May	2018
	5394	s5395	Movie	Hans Teeuwen: Real Rancour	Doesjka van Hoogdalem	Hans Teeuwen	Netherlands	2018	TV-MA	86 min	Stand-Up Comedy	Comedian Hans Teeuwen rebels against political...	July	2017
	5658	s5659	TV Show	Sense8	unknown	Tuppence Middleton, Brian J. Smith, Doona Bae,...	United states	2018	TV-MA	2 Seasons	Crime TV Shows, TV Dramas, TV Mysteries	From the creators of "The Matrix" and "Babylon...	December	2016
	5677	s5678	TV Show	Tokyo Trial	unknown	Marcel Hensema, Paul Freeman, Irrfan Khan, Sta...	Canada, japan, netherlands	2017	TV-MA	1 Season	International TV Shows, TV Dramas	In the wake of World War II, 11 Allied judges ...	December	2016
	7063	s7064	Movie	Incoming	Eric Zaragosa	Scott Adkins, Aaron McCusker, Vahidin Prelic, ...	Serbia, united states	2019	TV-MA	89 min	Action & Adventure, Sci-Fi & Fantasy	When an imprisoned terrorist cell hijacks a hi...	October	2018
	7112	s7113	TV Show	Jack Taylor	Stuart Orme	Iain Glen, Killian Scott, Nora-Jane Noone, Tar...	United states, ireland	2016	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Dramas	Ken Bruen's bestselling crime novels leap to l...	March	2013

In [31]:

df.head()

Out[31]:

	show_id	type	title	director	cast	country	release_year	rating	duration	listed_in	description	month_added	year_added
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	unknown	United states	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	September	2021
1	s2	TV Show	Blood & Water	unknown	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South africa	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	September	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Unknown	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...	September	2021
3	s4	TV Show	Jailbirds New Orleans	unknown	unknown	Unknown	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...	September	2021
4	s5	TV Show	Kota Factory	unknown	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...	September	2021

In [32]:

Year added can't less the release year hence dropping those column
df.drop(df[df['year_added']<df['release_year']].index,inplace=True)

In [33]:

df[df['year_added']<df['release_year']].sum()

Out[33]:

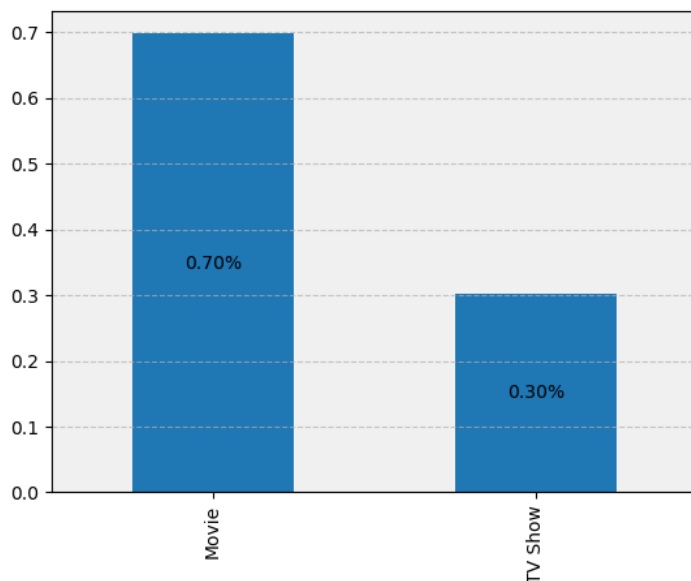
	show_id	type	title	director	cast	country	release_year	rating	duration	listed_in	description	month_added	year_added
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	unknown	United states	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	September	2021
1	s2	TV Show	Blood & Water	unknown	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South africa	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	September	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Unknown	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...	September	2021
3	s4	TV Show	Jailbirds New Orleans	unknown	unknown	Unknown	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...	September	2021
4	s5	TV Show	Kota Factory	unknown	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...	September	2021
...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United states	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...	November	2019
8803	s8804	TV Show	Zombie Dumb	unknown	unknown	Unknown	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...	July	2019
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United states	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...	November	2019
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United states	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...	January	2020
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	March	2019

8776 rows × 13 columns

```
In [34]: df['type'].value_counts()
```

```
Out[34]: Movie      6124
TV Show    2652
Name: type, dtype: int64
```

```
In [35]: ax=df['type'].value_counts(normalize=True).plot.bar()
ax.bar_label(ax.containers[0], fmt='%.2f%', label_type='center')
ax.grid(axis='y', linestyle='--', alpha= 0.7)
ax.set_facecolor('#f0f0f0')
```



On Netflix 70% are Movies & 30% are TV Shows

Analysing Director feature

```
In [36]: df['director'].nunique()
```

```
Out[36]: 4524
```

```
In [37]: df['director'].value_counts()
```

```
Out[37]: unknown      2610
Rajiv Chilaka      19
Raúl Campos, Jan Suter    18
Marcus Raboy      16
Suhas Kadav      16
...
Raymie Muzquiz, Stu Livingston    1
Joe Menendez      1
Eric Bross      1
Will Eisenberg    1
Mozes Singh      1
Name: director, Length: 4524, dtype: int64
```

- *Director Rajiv Chilaka has highest number of shows*

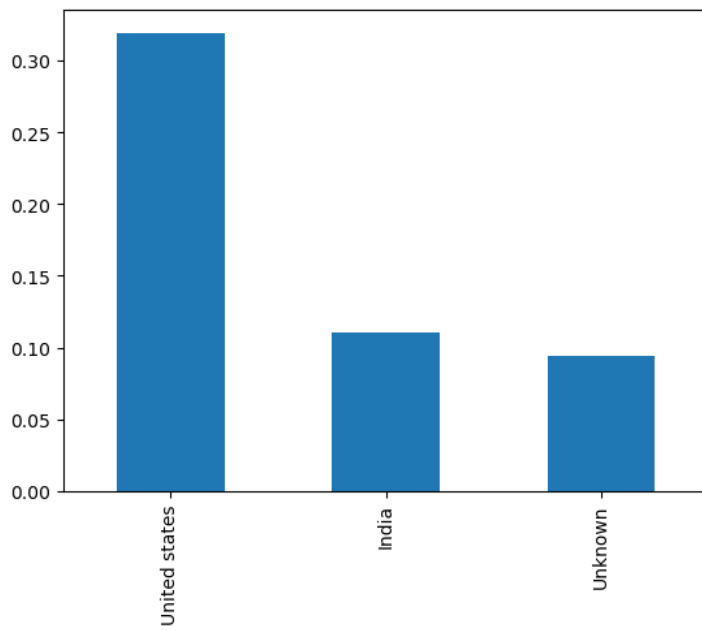
Analysing country feature

```
In [38]: df['country'].nunique()
```

```
Out[38]: 746
```

```
In [39]: df['country'].value_counts(normalize=True)[:3].plot.bar()
```

Out[39]: <Axes: >



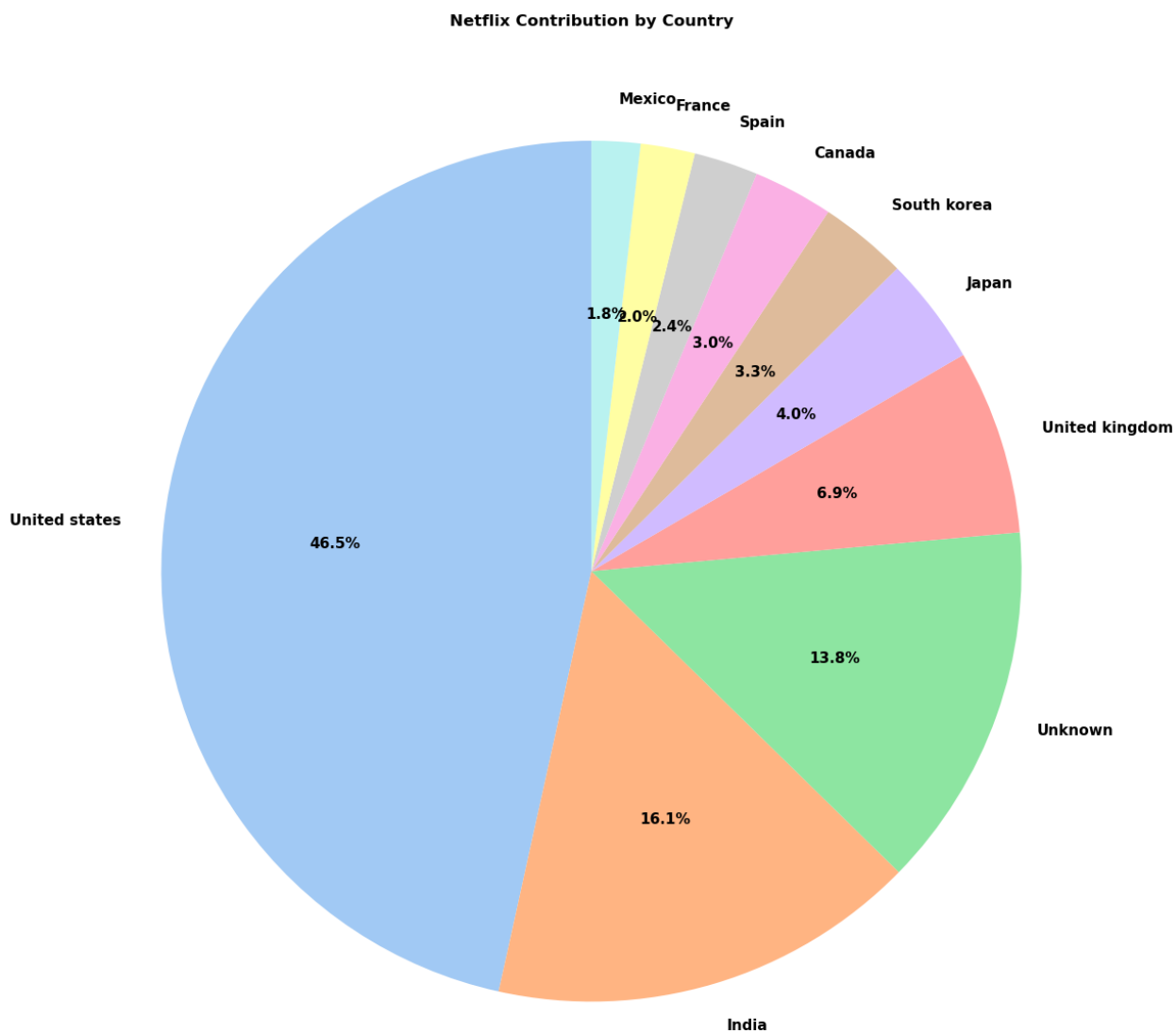
```
In [40]: country_counts = df['country'].value_counts(normalize=True)[:10]

# Create a pie chart using Matplotlib and Seaborn
plt.figure(figsize=(14.5, 14.5))
sns.set_palette("pastel") # Set color palette

# Plot the pie chart
plt.pie(country_counts, labels=country_counts.index, autopct='%1.1f%%', startangle=90,
        textprops={'fontsize':11,'fontweight':'bold'})

# Set plot title
plt.title('Netflix Contribution by Country',fontweight='heavy')

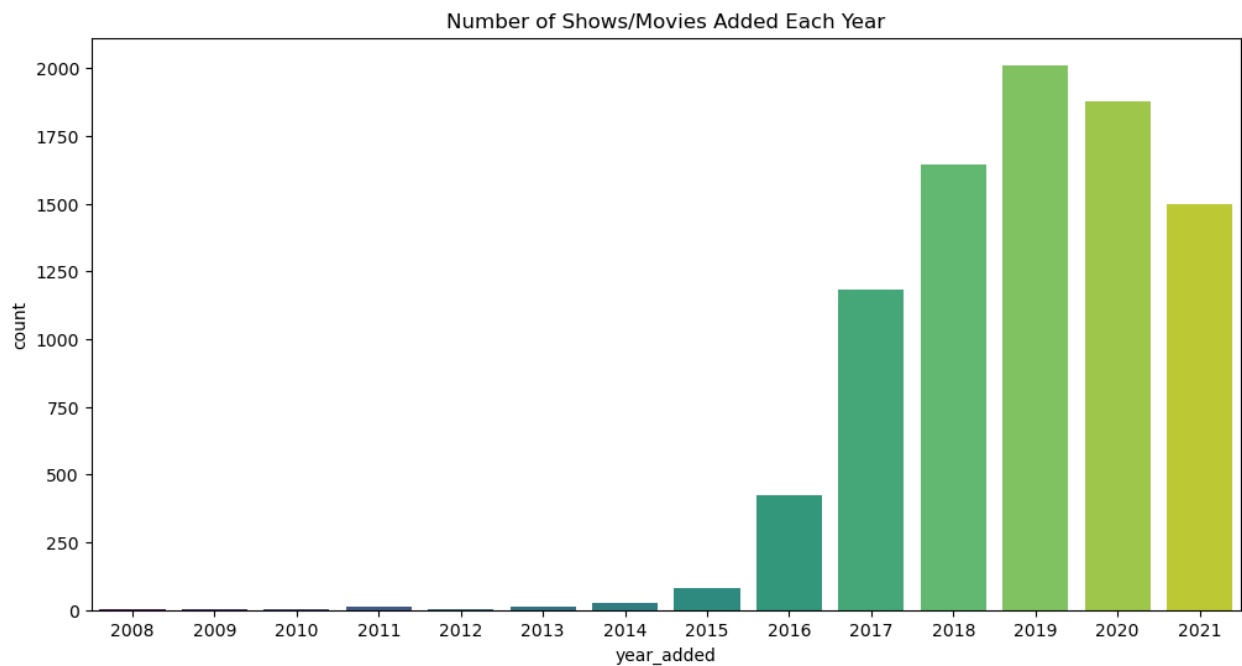
# Show the plot
plt.show()
```



- As per above chart United States & India contributes most in Netflix

Analysing Number of Shows/Movies Added Each Year

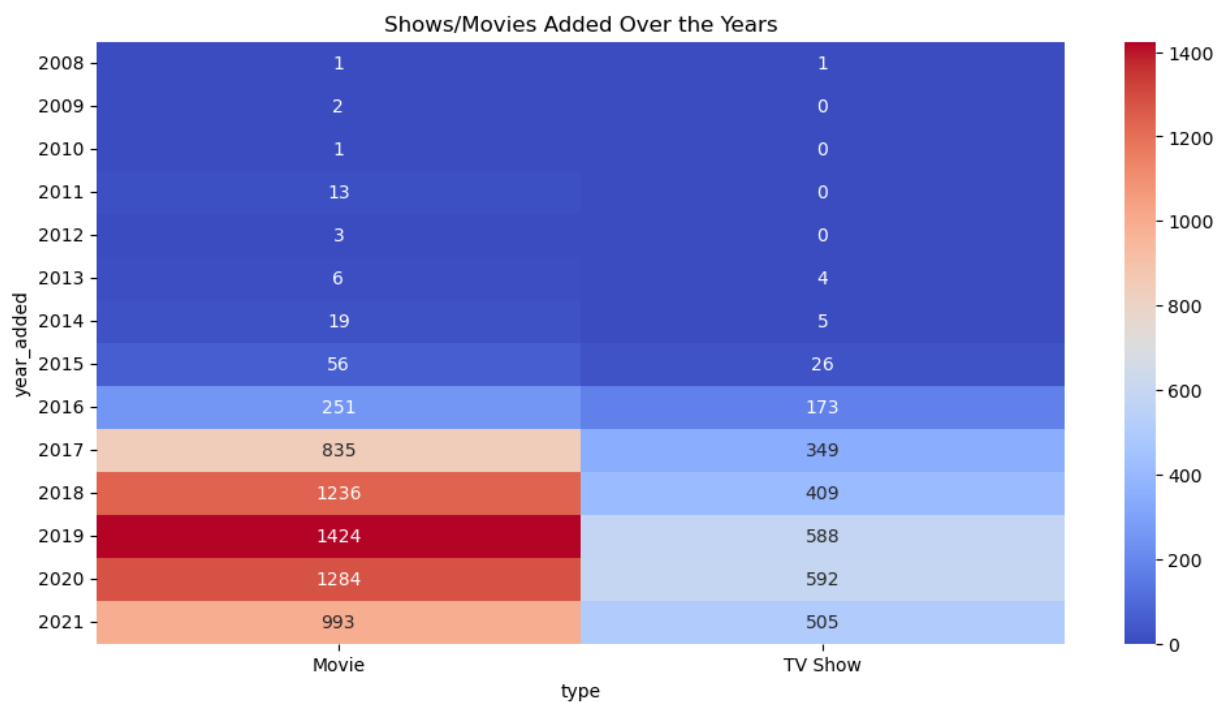
```
In [41]: plt.figure(figsize=(12, 6))
sns.countplot(df, x='year_added', palette='viridis')
plt.title('Number of Shows/Movies Added Each Year')
plt.show()
```



- 2019 gave the highest number of shows followed by 2020

Shows/Movies added over the Years

```
In [42]: plt.figure(figsize=(12, 6))
sns.heatmap(pd.crosstab(df['year_added'], df['type']), cmap='coolwarm', annot=True, fmt='d')
plt.title('Shows/Movies Added Over the Years')
plt.show()
```



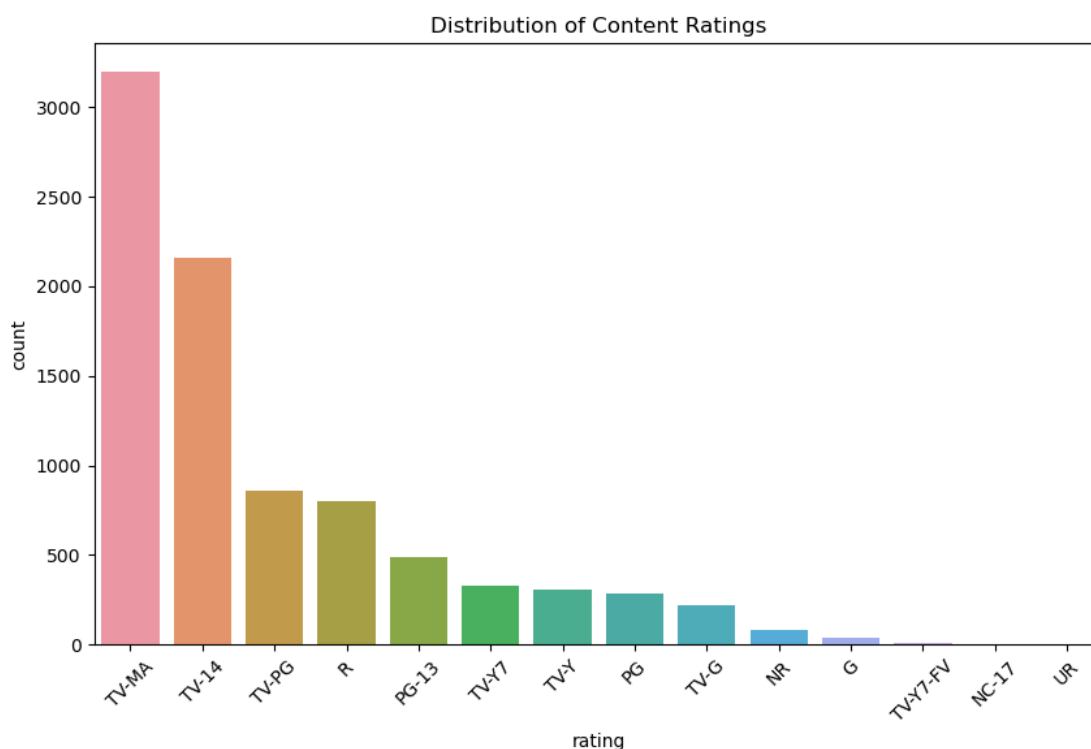
- On 2019 highest number of movies & 2020 highest number of TV shows added

```
In [43]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8776 entries, 0 to 8806
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype  
---  --
 0   show_id         8776 non-null   object  
 1   type            8776 non-null   object  
 2   title           8776 non-null   object  
 3   director        8776 non-null   object  
 4   cast            8776 non-null   object  
 5   country         8776 non-null   object  
 6   release_year    8776 non-null   int64   
 7   rating          8776 non-null   object  
 8   duration        8776 non-null   object  
 9   listed_in       8776 non-null   object  
10   description      8776 non-null   object  
11   month_added     8776 non-null   object  
12   year_added      8776 non-null   int32   
dtypes: int32(1), int64(1), object(11)
memory usage: 925.6+ KB
```

Analysing the content rating

```
In [44]: plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='rating', order=df['rating'].value_counts().index)
plt.title('Distribution of Content Ratings')
plt.xticks(rotation=45, ha='center')
plt.show()
```



- Most of the movies got rating TV-MA,TV-14,TV-PG

In [45]:

```
# Analysing Listed in feature
df.head()
```

Out[45]:

	show_id	type	title	director	cast	country	release_year	rating	duration	listed_in	description	month_added	year_added
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	unknown	United states	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	September	2021
1	s2	TV Show	Blood & Water	unknown	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South africa	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	September	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Unknown	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...	September	2021
3	s4	TV Show	Jailbirds New Orleans	unknown	unknown	Unknown	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...	September	2021
4	s5	TV Show	Kota Factory	unknown	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...	September	2021

In [46]:

```
# Analysing Listed_in

Split the 'listed_in' column into individual genres
genres_split = df['listed_in'].str.split(', ').explode()

# Count the occurrences of each genre
genre_counts = genres_split.value_counts()

# Create a pie chart
plt.figure(figsize=(10, 10))
genre_counts.head(10).plot.pie(autopct='%1.1f%%', startangle=90)
plt.title('Top 10 Genres Distribution on Netflix')
plt.show()
```

Cell In[46], line 3
Split the 'listed_in' column into individual genres

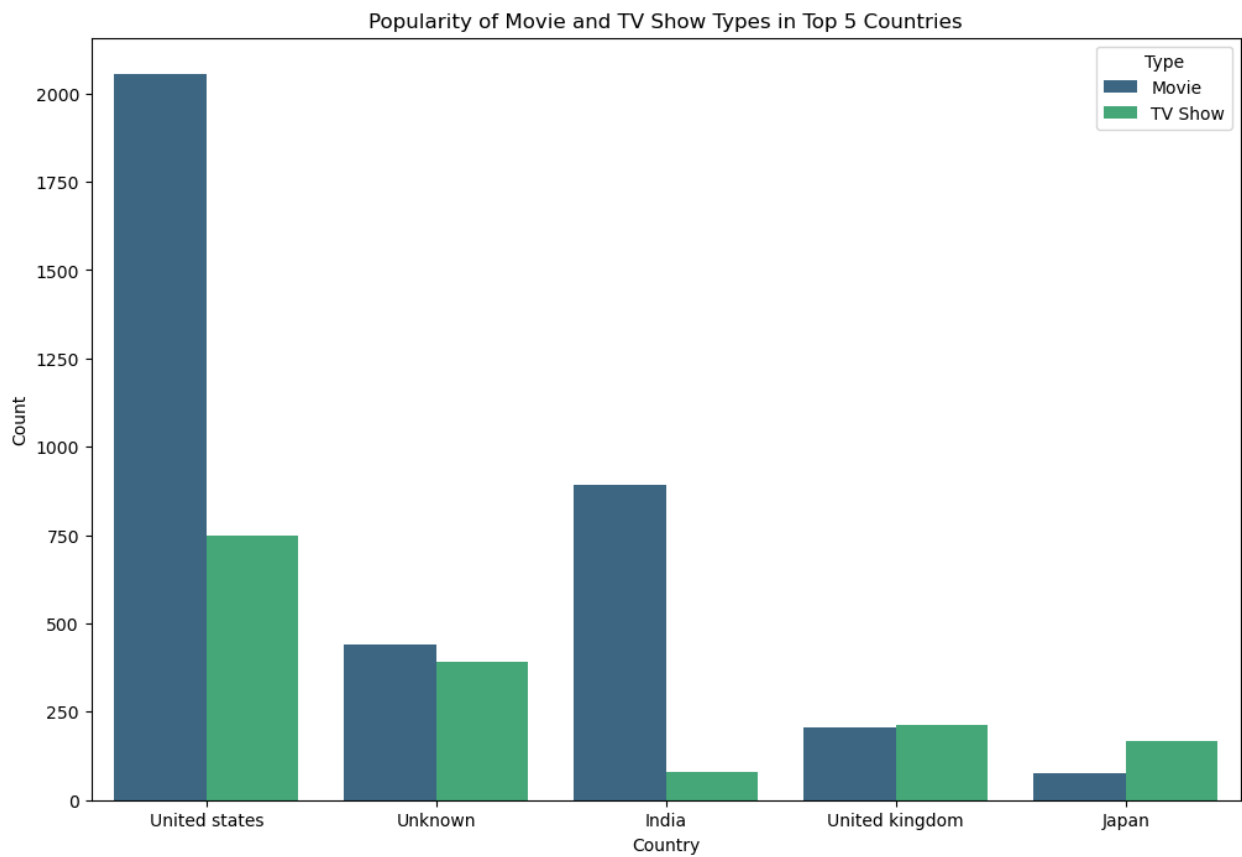
SyntaxError: invalid syntax

- Netflix has highest number of Internationala movies followed by Dramas and then Comedies.

Bivariate Analysis

```
In [47]: top_countries = df['country'].value_counts().head(5).index
temp_df = df[df['country'].isin(top_countries)]

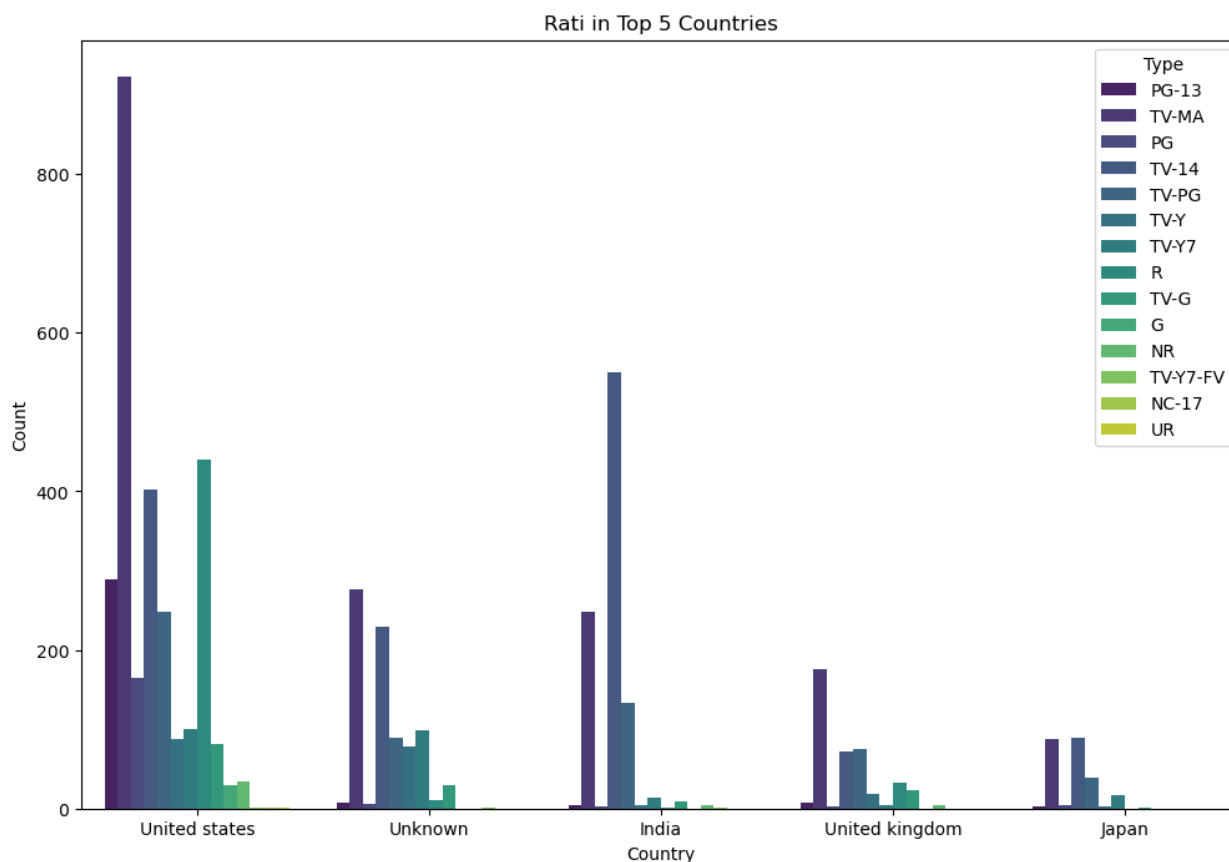
# Created a stacked bar plot
plt.figure(figsize=(12, 8))
sns.countplot(data=temp_df, x='country', hue='type', palette='viridis')
plt.title('Popularity of Movie and TV Show Types in Top 5 Countries')
plt.xlabel('Country')
plt.ylabel('Count')
plt.legend(title='Type', loc='upper right')
plt.show()
```



- In USA and India 'Movie' show are more famous than 'TV Shows', and in United Kingdom both type are equally popular, and in Japan 'TV shows' are more popular than 'Movie' shows.


```
In [48]: top_countries = df['country'].value_counts().head(5).index
temp_df = df[df['country'].isin(top_countries)]

# Created a stacked bar plot
plt.figure(figsize=(12, 8))
sns.countplot(data=temp_df, x='country', hue='rating', palette='viridis')
plt.title('Rati in Top 5 Countries')
plt.xlabel('Country')
plt.ylabel('Count')
plt.legend(title='Type', loc='upper right')
plt.show()
```



- For USA most of the shows are TV-MA rating and for India most of the shows has TV-14 rating

Summary

- On Netflix 70% are Movie and 30% are TV Show
- Director "Rajiv Chilaka" has released highest number of shows.
- On Netflix most of the shows are from USA(46.1%) and India(16.1%).
- Highest number of shows added was in 2019.
- On 2019 highest number of movies & 2020 highest number of TV shows added
- Most of the shows are of rating 'TV-MA,TV-14,TV-PG'.
- Netflix has highest (21.7%) shows of 'International Movies' genre,followed by 'Dramas' genre.
- In USA and India 'Movie' show are more famous than 'TV Shows',and in United Kingdom both type are equily popular,and In Japan 'TV shows' are more popular than 'Movie' shows.
- for United states most of the shows have 'TV-MA' rating and in India most of the shows have 'TV-14' rating.

In []: