

+ Code

+ Text

```
#1and2. load the file and print 5 rows of data
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
path="/content/netflix_titles.csv.zip"
df=pd.read_csv(path)
df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV	To protect his family from a powerful

Next steps:

[Generate code with df](#)[View recommended plots](#)[New interactive sheet](#)

```
#3 print last 5 rows of the data
df.tail()
```


	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...

```
#4 Identifyinf missing values
df.isnull()
#df.dropna().sum()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	False	False	False	False	True	False	False	False	False	False	False	False
1	False	False	False	True	False	False	False	False	False	False	False	False
2	False	False	False	False	False	True	False	False	False	False	False	False
3	False	False	False	True	True	True	False	False	False	False	False	False
4	False	False	False	True	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	False	False	False	False	False	False	False	False	False	False	False	False
8803	False	False	False	True	True	True	False	False	False	False	False	False
8804	False	False	False	False	False	False	False	False	False	False	False	False
8805	False	False	False	False	False	False	False	False	False	False	False	False
8806	False	False	False	False	False	False	False	False	False	False	False	False

8807 rows × 12 columns

```
#5 Cleaning the null values
df.dropna()
```



	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descriptio
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies	On a phot shoot Ghana, a America model s.
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...	United Kingdom	September 24, 2021	2021	TV-14	9 Seasons	British TV Shows, Reality TV	A talente batch ( amateu bakers fac off in.
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG-13	104 min	Comedies, Dramas	A woma adjusting t life after los contend.
12	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	2021	TV-MA	127 min	Dramas, International Movies	After most ( her family murdered a terr.
24	s25	Movie	Jeans	S. Shankar	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...	India	September 21, 2021	1998	TV-14	166 min	Comedies, International Movies, Romantic Movies	When th father of th man sh loves insisi t.
...	...	...	...	...	...	...	...	...	...	...	...	...
8801	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Recoverin alcohol Talal wake up inside s.
					Mark Ruffalo,						Cult Movies	A politic cartoonist

```
#5 cleaning the null values
df.dropna()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafulmike Ogunlana, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies	On a phot shoot Ghana, a America model s.
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...	United Kingdom	September 24, 2021	2021	TV-14	9 Seasons	British TV Shows, Reality TV	A talente batch ( amate bakers fac off in.
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG-13	104 min	Comedies, Dramas	A woma adjusting l life after los contend.
12	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	2021	TV-MA	127 min	Dramas, International Movies	After most ( her family murdered a terr.
24	s25	Movie	Jeans	S. Shankar	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...	India	September 21, 2021	1998	TV-14	166 min	Comedies, International Movies, Romantic Movies	When th father of th man sh loves insis t.
...	...	...	...	...	...	...	...	...	...	...	...	...
8801	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Recoverin alcohol Talal wake up inside s.
					Mark Ruffalo,						Cult Movies	A politic cartoonist


#6 info on the dataset  
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


```

#7 Get some description of the data-release\_year  
df.describe()




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

#8 columns on the dataset  
df.columns



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

#9 shape of the data frame  
df.shape




```
(8807, 12)
```

#10 how many years of data ( release\_year)  
df['release\_year'].count()



```
8807
```

#11 Movies Vs TV shows (type)  
movies\_df=df[df['type']=='Movie']  
tv\_shows\_df=df[df['type']=='TV Show']  
print(movies\_df.dtypes)  
print(tv\_shows\_df.dtypes)



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

#12 Check for null values in all columns of dataset  
df.isnull().sum()



0

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

```
#13 Replace NAN values with TV-MA
df.fillna('United States')
```



	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	United States	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	United States	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	United States	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	United States	United States	United States	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	United States	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and

```
#14 Replace NAN values with United States
df.fillna('united States')
```

	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	united States	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	united States	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	united States	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	united States	united States	united States	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	united States	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and

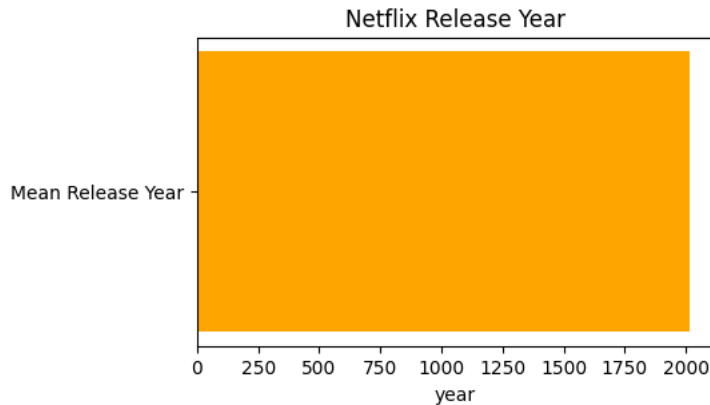
```
#15 Rename the column " listed_in"to"genre"
df.rename(columns={'listed_in':'genre'}, inplace=True)
print(df.head())
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	genre	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	united States	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	united States	united States	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...

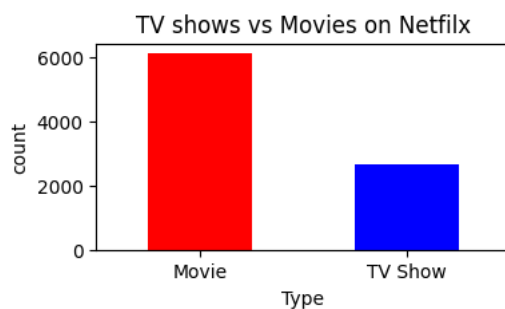
```
#VISUALIZATION#
```

```
mean_release_year=df['release_year'].mean()
plt.figure(figsize=(5,3))
```

```
plt.barh(['Mean Release Year'],[mean_release_year],color='orange')
plt.xlabel('year')
plt.title('Netflix Release Year')
plt.show()
```



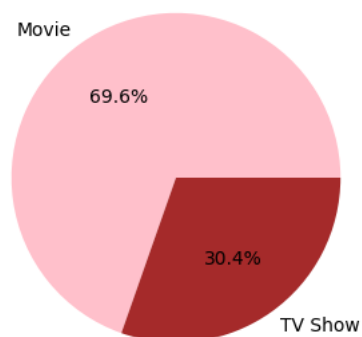
```
type_counts=df['type'].value_counts()
plt.figure(figsize=(4,2))
type_counts.plot(kind='bar',color=['red', 'blue'])
plt.xlabel('Type')
plt.ylabel('count')
plt.title('TV shows vs Movies on Netfilx')
plt.xticks(rotation=0)
plt.show()
```



```
type_counts=df['type'].value_counts()
plt.figure(figsize=(4,4))
plt.pie(type_counts, labels=type_counts.index, autopct='%1.1f%%',
        colors=['pink','brown'])
plt.title('Percentage of show Types on Netflix')
plt.show()
```



Percentage of show Types on Netflix

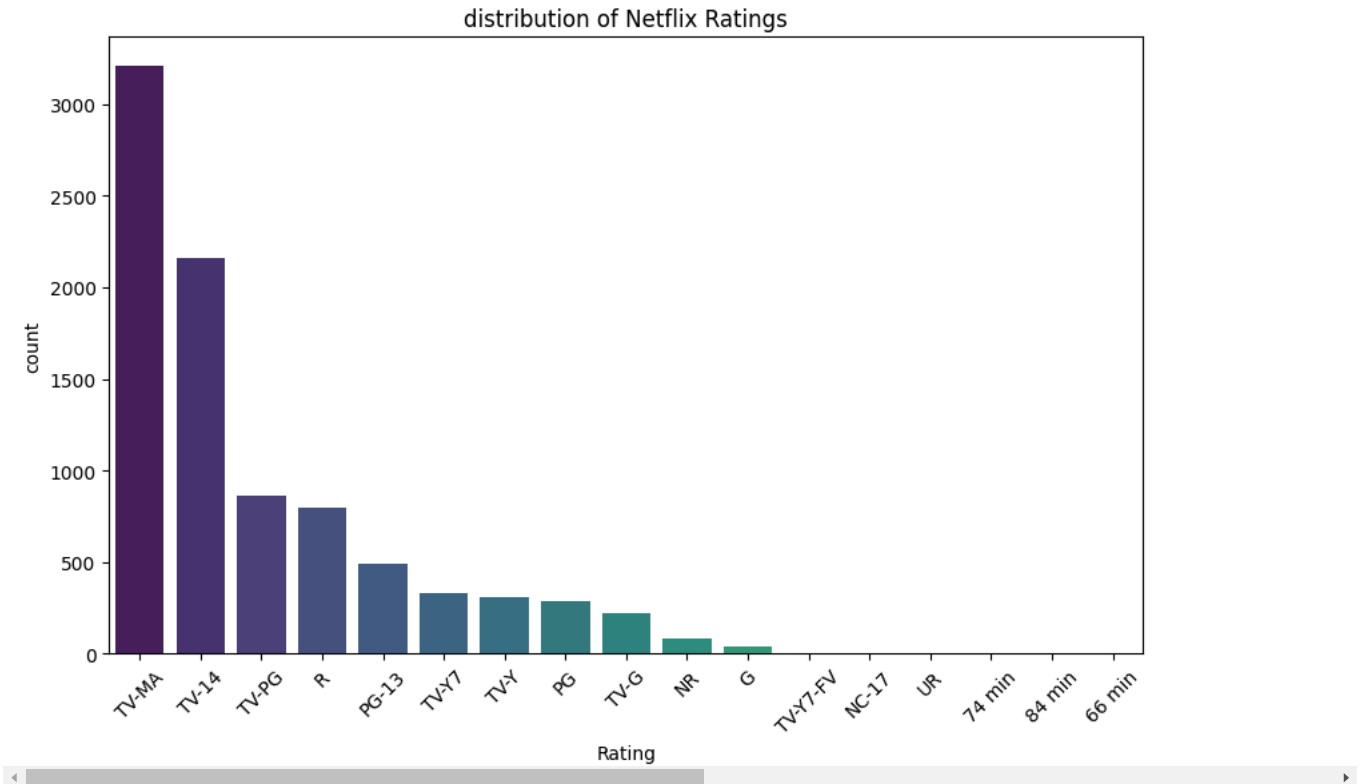


```
plt.figure(figsize=(10,6))
sns.countplot(data=df, x='rating', order=df['rating'].value_counts().index,palette='viridis')
plt.xlabel('Rating')
plt.ylabel('count')
plt.title('distribution of Netflix Ratings')
plt.xticks(rotation=45)
plt.show()
```

```
<ipython-input-22-c809c9bf65c6>:2: FutureWarning:
```

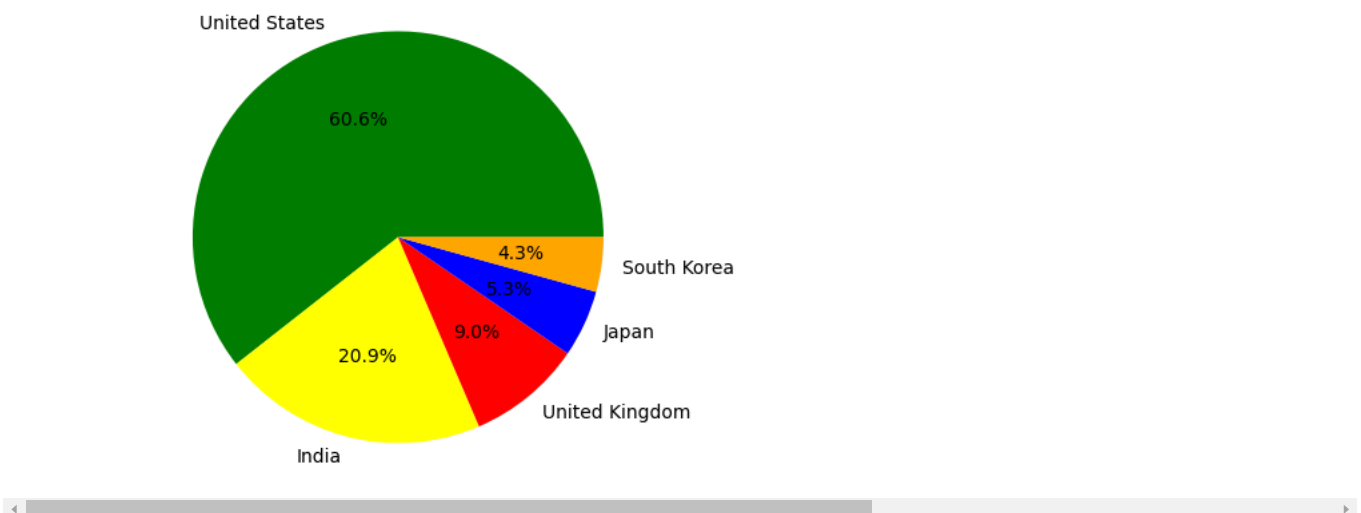
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `le

```
sns.countplot(data=df, x='rating', order=df['rating'].value_counts().index,palette='viridis')
```



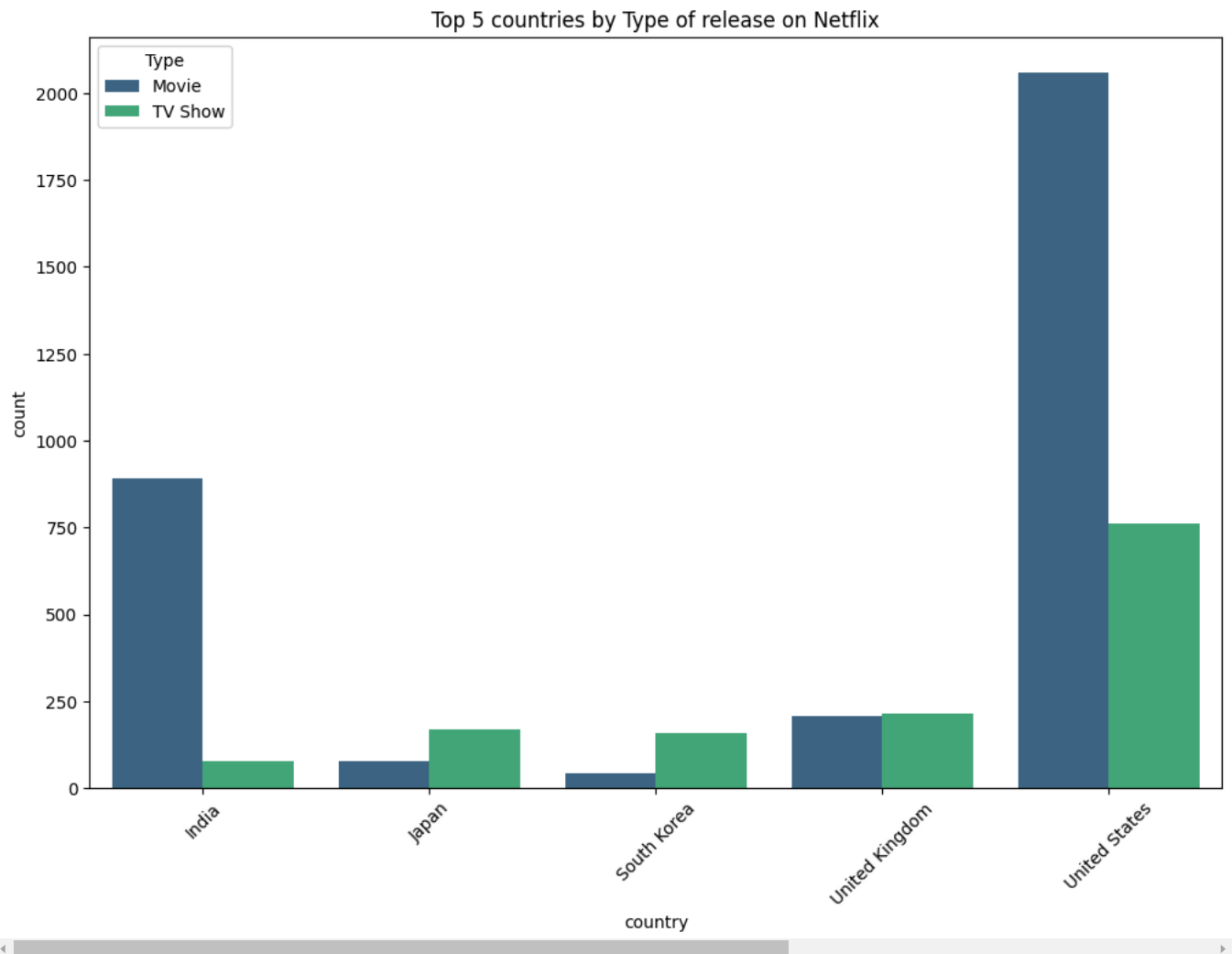
```
country_counts=df['country'].value_counts()
top_5_countries=country_counts.head(5)
plt.figure(figsize=(5,5))
plt.pie(top_5_countries, labels=top_5_countries.index, autopct='%1.1f%%',
colors=['green','yellow','red','blue','orange'])
plt.title('Top 5 countries with Highest Number of Movies/TV shows on Netfilx')
plt.show()
```

**Top 5 countries with Highest Number of Movies/TV shows on Netflix**

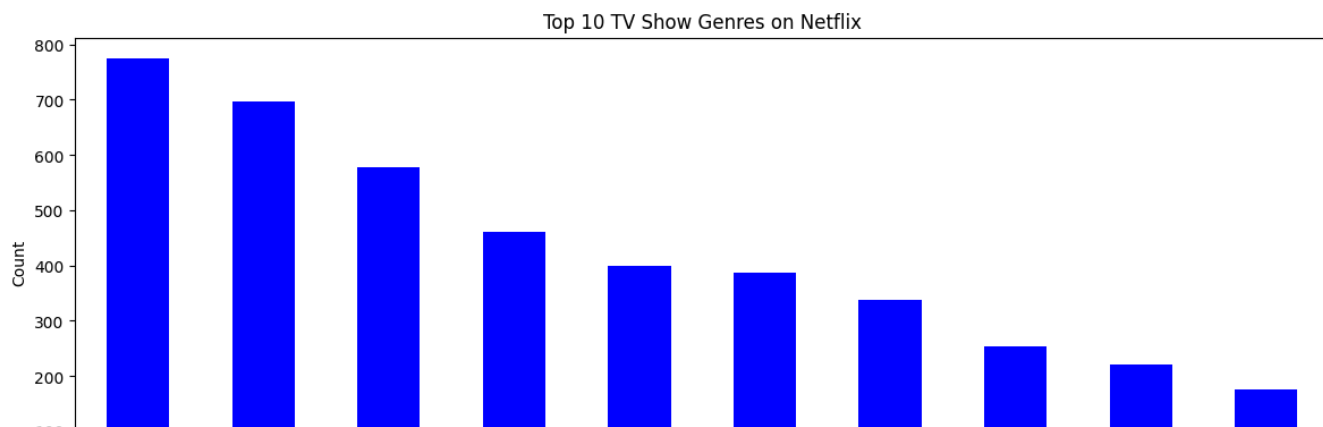
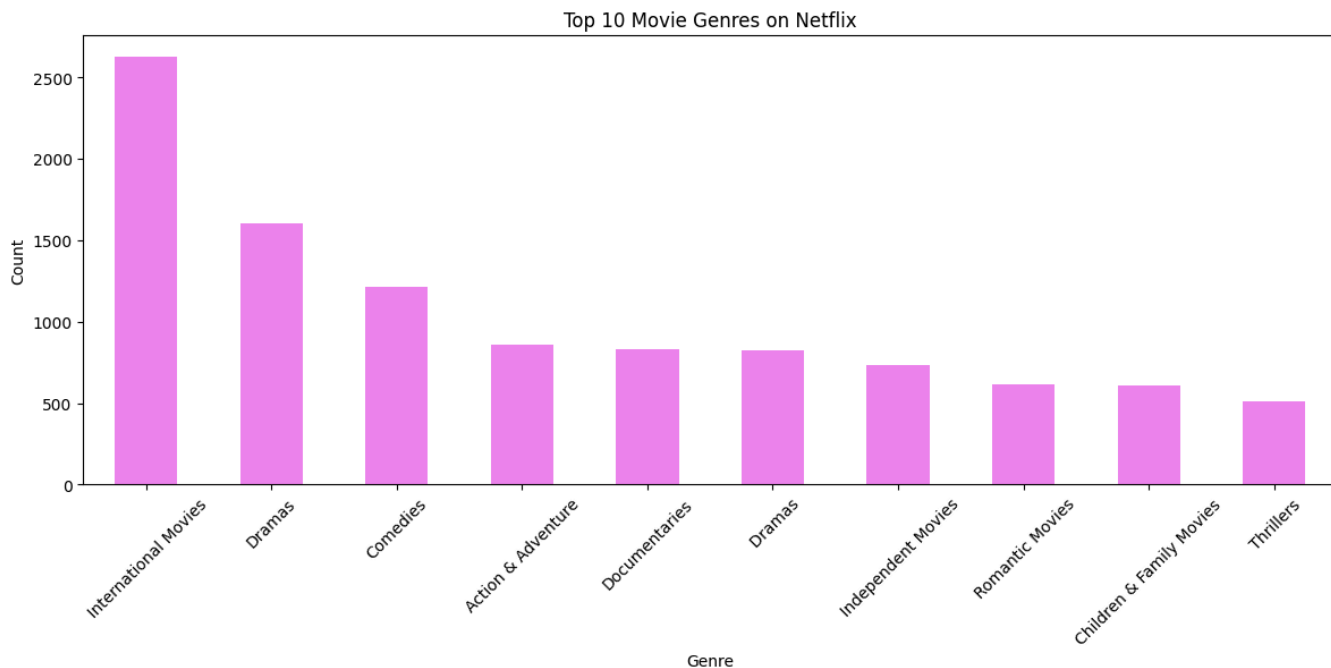


```
country_type_counts=df.groupby(['country', 'type']).size().reset_index(name='count')
top_countries=country_type_counts.groupby('country')['count'].sum().nlargest(5).index
filtered_data=country_type_counts[country_type_counts['country'].isin(top_countries)]
plt.figure(figsize=(12,8))
sns.barplot(data=filtered_data, x='country',y='count',hue='type',palette='viridis')
plt.xlabel('country')
plt.ylabel('count')
plt.title('Top 5 countries by Type of release on Netflix')
plt.xticks(rotation=45)
plt.legend(title='Type')
plt.show()
```





```
df['genre'] = df['genre'].str.split(',')
df = df.explode('genre')
movies_df = df[df['type'] == 'Movie']
tv_shows_df = df[df['type'] == 'TV Show']
top_10_movie_genres = movies_df['genre'].value_counts().head(10)
top_10_tv_genres = tv_shows_df['genre'].value_counts().head(10)
fig, axes = plt.subplots(2, 1, figsize=(12, 12))
top_10_movie_genres.plot(kind='bar', ax=axes[0], color='violet')
axes[0].set_title('Top 10 Movie Genres on Netflix')
axes[0].set_xlabel('Genre')
axes[0].set_ylabel('Count')
axes[0].set_xticklabels(top_10_movie_genres.index, rotation=45)
top_10_tv_genres.plot(kind='bar', ax=axes[1], color='blue')
axes[1].set_title('Top 10 TV Show Genres on Netflix')
axes[1].set_xlabel('Genre')
axes[1].set_ylabel('Count')
axes[1].set_xticklabels(top_10_tv_genres.index, rotation=45)
plt.tight_layout()
plt.show()
```



```
df['date_added']=pd.to_datetime(df['date_added'],errors='coerce')
df['Month']=df['date_added'].dt.month # Changed Month to month
df['DayOfWeek']=df['date_added'].dt.day_name() # Changed data to df
release_counts=df.groupby(['Month','DayOfWeek']).size().unstack() # Changed data to df
print("Aggregated release counts by month and day of the week:")
print(release_counts)
plt.figure(figsize=(12,8))
sns.heatmap(release_counts, annot=True, cmap="YlGnBu", fmt="d", cbar=True,
linewidths=0.5)
plt.title('Number of Release by month and Day of the Week')
plt.xlabel('Day of the Week')
plt.ylabel('Month')
plt.xticks(rotation=45)
plt.yticks(rotation=0)
plt.show()
```



Aggregated release counts by month and day of the week:

DayOfWeek	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday	
Month	1.0	425	217	70	100	138	218	369

