

importing python libraries

```
In [116]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
In [34]: # pd.set_option('display.max_columns', None) # Show all columns  
# pd.set_option('display.width', 1000) # Set a larger width to avoid wrapping  
# pd.set_option('display.max_colwidth', 50) # Disable column width truncation
```

Importing dataset

```
In [155]: df = pd.read_csv(r'D:\Learning\scaler data science\python adv\business case\Netflix.txt', delimiter=',')  
print(df.head())
```

show_id	type	title	director	listed_in	cast	country	date_added	release_year	rating	duration	
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	As her father nears the end of his life, filmm...	NaN	United States	September 25, 2021	2020	PG-13	90 mi
1	s2	TV Show	Blood & Water	Nan Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Season		
2	s3	TV Show	Ganglands	Julien Leclercq Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Seaso		
3	s4	TV Show	Jailbirds New Orleans	Nan	NaN	September 24, 2021	2021	TV-MA	1 Seaso		
4	s5	TV Show	Kota Factory	Nan Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Season		
s				In a city of coaching centers known to train I...							

```
In [71]: print(df.columns)  
print(df.shape)  
# print(df['type'])
```

Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added', 'release_year', 'rating', 'duration', 'listed_in', 'description'], dtype='object')
(8807, 12)

0	Movie
1	TV Show
2	TV Show
3	TV Show
4	TV Show
...	
8802	Movie
8803	TV Show
8804	Movie
8805	Movie
8806	Movie

Name: type, Length: 8807, dtype: object

```
In [37]: print(df.info())  
print(df.describe())
```

```
<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
None
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 [110]: print(df.isnull().sum())
print(df.unique())
# print(df['country'].unique())
```

```
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
year_release  10
dtype: int64
show_id      8807
type         2
title        8807
director    4528
cast         7692
country      748
date_added   1767
release_year  74
rating        17
duration      220
listed_in     514
description    8775
year_release  14
dtype: int64
```

```
In [39]: columns = df.columns
for column in columns:
    print(f'{column}: ', df[column].value_counts())
```

```
show_id: show_id
s1      1
s5875   1
s5869   1
s5870   1
s5871   1
..
s2931   1
s2930   1
s2929   1
s2928   1
s8807   1
Name: count, Length: 8807, dtype: int64
type: type
Movie      6131
TV Show    2676
Name: count, dtype: int64
title: title
Dick Johnson Is Dead          1
Ip Man 2                      1
Hannibal Buress: Comedy Camisado 1
Turbo FAST                     1
Masha's Tales                  1
..
Love for Sale 2                1
ROAD TO ROMA                   1
Good Time                      1
Captain Underpants Epic Choice-o-Rama 1
Zubaan                           1
Name: count, Length: 8807, dtype: int64
director: director
Rajiv Chilaka                  19
RaÃ±el Campos, Jan Suter       18
Marcus Raboy                   16
Suhas Kadav                    16
Jay Karas                      14
..
Raymie Muzquiz, Stu Livingston 1
Joe Menendez                   1
Eric Bross                      1
Will Eisenberg                 1
Mozez Singh                     1
Name: count, Length: 4528, dtype: int64
cast: cast
David Attenborough              19
Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jigna Bhardwaj, Rajesh Kava, Mousam, Swapnil 14
Samuel West                      10
Jeff Dunham                      7
David Spade, London Hughes, Fortune Feimster 6
..
Michael PeÃ±ta, Diego Luna, Tenoch Huerta, Joaquin Cosio, JosÃ© MarÃ¡a Yazpik, Matt Letscher, Alyssa Diaz 1
Nick Lachey, Vanessa Lachey     1
```

Takeru Sato, Kasumi Arimura, Haru, Kentaro Sakaguchi, Takayuki Yamada, Kendo Kobayashi, Ken Yasuda, Arata Furuta, Suzuki Matsuo, Koichi Yamadera, Arata Iura, Chikako Kaku, Kot aro Yoshida 1
Toyin Abraham, Sambasa Nzeribe, Chioma Chukwuka Akpotta, Chioma Omeruah, Chiwetalu Agu, Dele Odule, Femi Adebayo, Bayray McNwizu, Biodun Stephen
1
Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary, Meghna Malik, Malkeet Rauni, Anita Shabdish, Chittaranjan Tripathy
1
Name: count, Length: 7692, dtype: int64
country: country
United States 2818
India 972
United Kingdom 419
Japan 245
South Korea 199
...
Romania, Bulgaria, Hungary 1
Uruguay, Guatemala 1
France, Senegal, Belgium 1
Mexico, United States, Spain, Colombia 1
United Arab Emirates, Jordan 1
Name: count, Length: 748, dtype: int64
date_added: date_added
January 1, 2020 109
November 1, 2019 89
March 1, 2018 75
December 31, 2019 74
October 1, 2018 71
...
December 4, 2016 1
November 21, 2016 1
November 19, 2016 1
November 17, 2016 1
January 11, 2020 1
Name: count, Length: 1767, dtype: int64
release_year: release_year
2018 1147
2017 1032
2019 1030
2020 953
2016 902
...
1959 1
1925 1
1961 1
1947 1
1966 1
Name: count, Length: 74, dtype: int64
rating: rating
TV-MA 3207
TV-14 2160
TV-PG 863
R 799
PG-13 490
TV-Y7 334
TV-Y 307
PG 287
TV-G 220
NR 80
G 41
TV-Y7-FV 6
NC-17 3

```

UR          3
74 min      1
84 min      1
66 min      1
Name: count, dtype: int64
duration: duration
1 Season    1793
2 Seasons   425
3 Seasons   199
90 min     152
94 min     146
...
16 min      1
186 min     1
193 min     1
189 min     1
191 min     1
Name: count, Length: 220, dtype: int64
listed_in: listed_in
Dramas, International Movies           362
Documentaries                          359
Stand-Up Comedy                        334
Comedies, Dramas, International Movies 274
Dramas, Independent Movies, International Movies 252
...
Kids' TV, TV Action & Adventure, TV Dramas      1
TV Comedies, TV Dramas, TV Horror            1
Children & Family Movies, Comedies, LGBTQ Movies 1
Kids' TV, Spanish-Language TV Shows, Teen TV Shows 1
Cult Movies, Dramas, Thrillers                1
Name: count, Length: 514, dtype: int64
description: description
Paranormal activity at a lush, abandoned property alarms a group eager to redevelop the site, but the eerie events may not be as unearthly as they think.          4
Challenged to compose 100 songs before he can marry the girl he loves, a tortured but passionate singer-songwriter embarks on a poignant musical journey.          3
A surly septuagenarian gets another chance at her 20s after having her photo snapped at a studio that magically takes 50 years off her life.          3
Multiple women report their husbands as missing but when it appears they are looking for the same man, a police officer traces their cryptic connection. 3
Secrets bubble to the surface after a sensual encounter and an unforeseen crime entangle two friends and a woman caught between them.          2
..
Sent away to evade an arranged marriage, a 14-year-old begins a harrowing journey of sex work and poverty in the slums of Accra.          1
When his partner in crime goes missing, a small-time crook's life is transformed as he dedicates himself to raising the daughter his friend left behind. 1
During 1962's Cuban missile crisis, a troubled math genius finds himself drafted to play in a U.S.-Soviet chess match and a deadly game of espionage. 1
A teen's discovery of a vintage Polaroid camera develops into a darker tale when she finds that whoever takes their photo with it dies soon afterward. 1
A scrappy but poor boy worms his way into a tycoon's dysfunctional family, while facing his fear of music and the truth about his past.          1
Name: count, Length: 8775, dtype: int64

```

I. Comparision between movies and TV shows in different countries

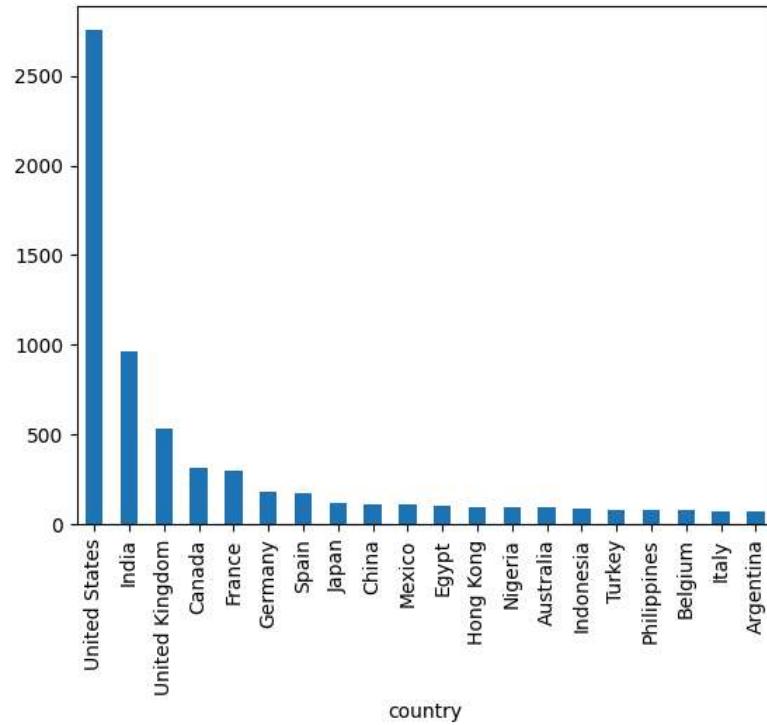
```
In [96]: df_movie = df[df['type'] == 'Movie']
print(df_movie['show_id'].count())
df_movie.loc[:, 'country'] = df_movie['country'].dropna().apply(lambda x : [i.strip() for i in x.split(',')])
df_movie = df_movie.explode('country')[['country']].value_counts()
print(df_movie)
print(df_movie.count())
df_movie.head(20).plot(kind='bar')
```

```

6131
country
United States    2752
India            962
United Kingdom   534
Canada           319
France           303
...
Bermuda          1
Ecuador          1
Armenia          1
Mongolia         1
Montenegro      1
Name: count, Length: 118, dtype: int64
118

```

```
Out[96]: <Axes: xlabel='country'>
```



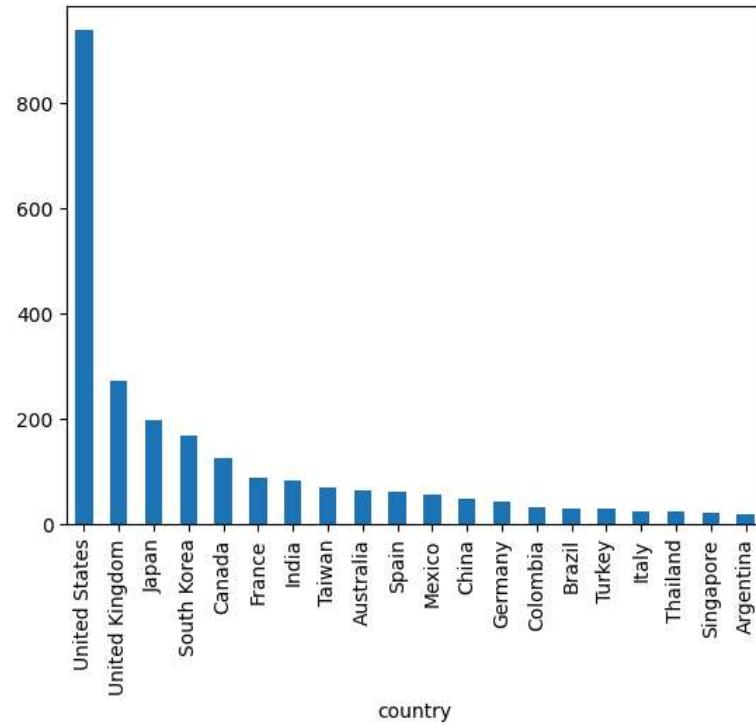
```
In [95]: df_tv_show = df[df['type'] == 'TV Show']
print(df_tv_show['show_id'].count())
df_tv_show.loc[:, 'country'] = df_tv_show['country'].dropna().apply(lambda x : [i.strip() for i in x.split(',')])
df_tv_show = df_tv_show.explode('country')['country'].value_counts()
print(df_tv_show)
print(df_tv_show.count())
df_tv_show.head(20).plot(kind='bar')
```

```

2676
country
United States      938
United Kingdom     272
Japan              199
South Korea        170
Canada             126
...
Malta              1
Belarus             1
United Arab Emirates 1
Uruguay             1
Switzerland         1
Name: count, Length: 66, dtype: int64
66

```

Out[95]: <Axes: xlabel='country'>



We can clearly see the movies vs TV show results as per the data set. There are total of 6171 movies and 2676 Tv shows on netflix.

Few of the business insights from the above data:

- 1 - United states produces max content with respect to both movies and TV shows.
- 2 - Netflix has content from total of 118 countries in movies and 66 countries in TV shows.
- 3 - India produces more movies as compared to TV shows.

Few of the recommendations from the above data:

1 - Netflix should also focus on countries whose content are very less on Netflix. It will increase the reach of netflix in more countries.

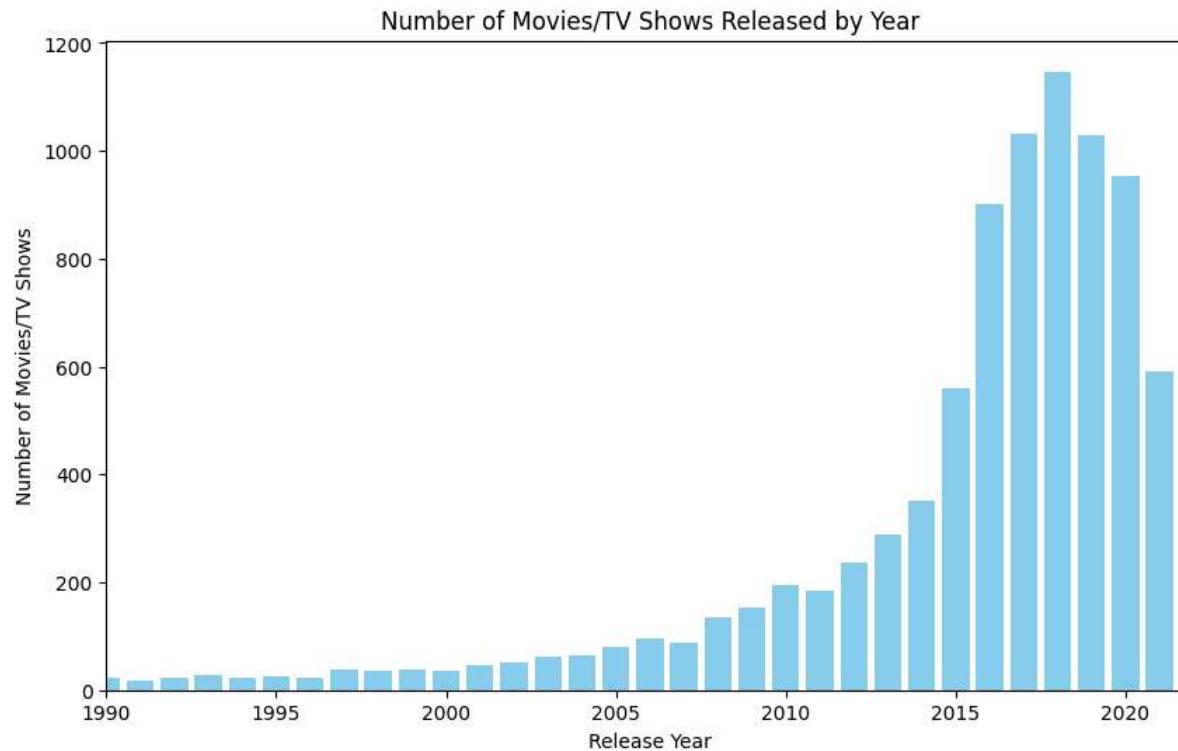
```
#####
```

II. Trends of movies/TV shows released every year.

```
In [ ]: #df.head()
df_year = df['release_year'].value_counts().sort_index()
df_year
```

```
Out[ ]: release_year
1925      1
1942      2
1943      3
1944      3
1945      4
...
2017    1032
2018    1147
2019    1030
2020     953
2021     592
Name: count, Length: 74, dtype: int64
```

```
In [133... plt.figure(figsize=(10, 6))
plt.bar(df_year.index, df_year.values, color='skyblue')
plt.xlabel('Release Year')
plt.ylabel('Number of Movies/TV Shows')
plt.title('Number of Movies/TV Shows Released by Year')
plt.xlim(1990, 2022)
plt.show()
```



As compared to last 20 years the number of movies/TV Shows produced has increased more than 30 times.

Maximum number of movies/TV Shows in netflix is from year 2018.

Business Insights : people are more interested in recent movies then old movies. SO netflix should focus more on recent releases then uploading old contents.

#####

III. filter ouf top 5 countries and top 5 genres in netflix. now see the trend whcih country produce which genre of movies more.

```
In [248]: df = pd.read_csv(r'D:\Learning\scaler data science\python adv\business case\Netflix.txt', delimiter=',')
print(df.head())

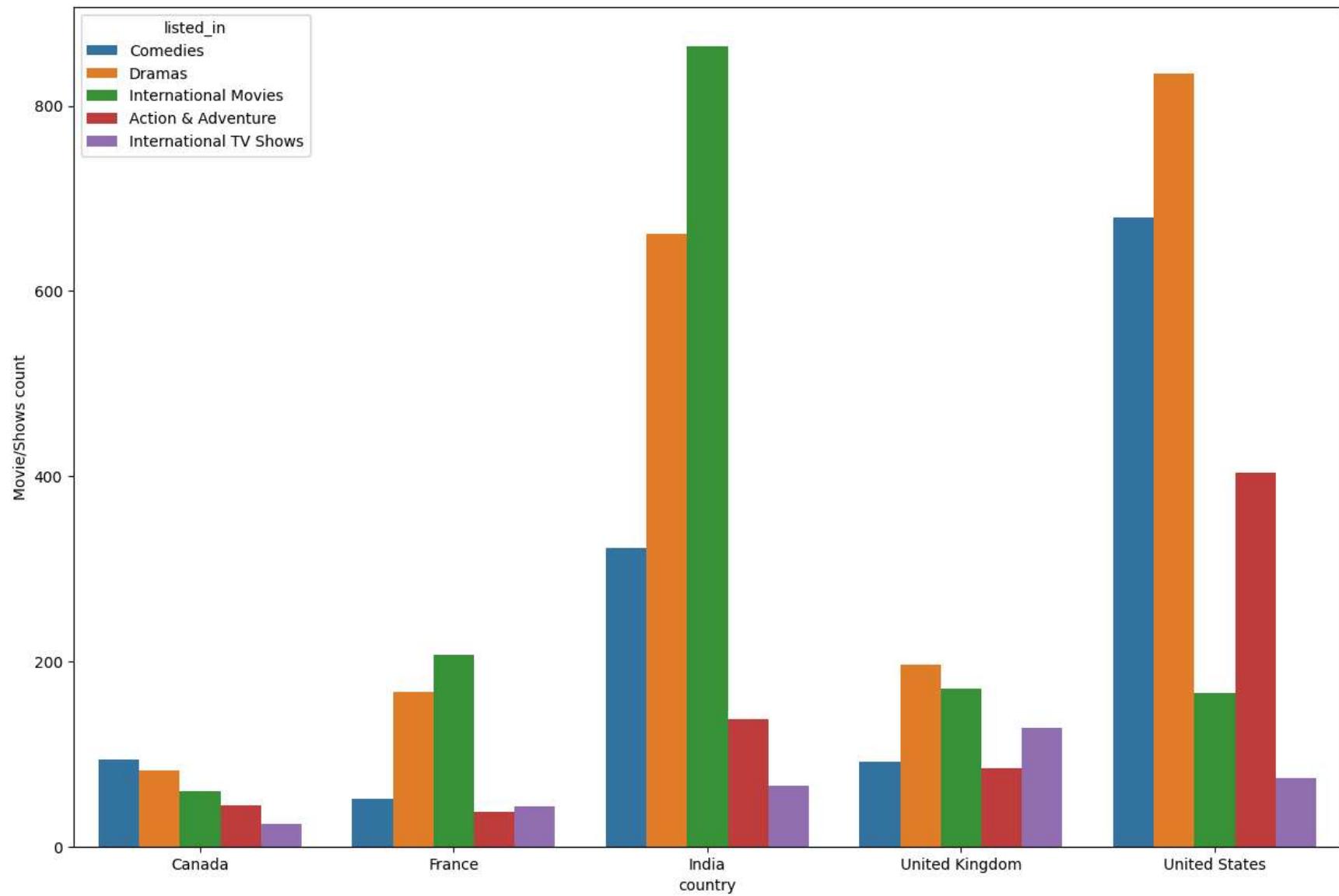
```

	show_id	type	title	director	listed_in	description	cast	country	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Documentaries	As her father nears the end of his life, filmm...	NaN	United States	September 25, 2021	2020	PG-13	90 mi
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Season		
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Seaso		
3	s4	TV Show	Jailbirds New Orleans	NaN	Feuds, flirtations and toilet talk go down amo...	NaN	September 24, 2021	2021	TV-MA	1 Seaso		
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Season		
5	s6	International TV Shows, TV Dramas, TV Mystery...	After crossing paths at a party, a Cape Town t...	NaN	To protect his family from a powerful drug lor...	NaN	September 24, 2021	2021	TV-MA	1 Seaso		
6	s7	Crime TV Shows, International TV Shows, TV Act...	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Seaso		
7	s8	TV Show	Jailbirds New Orleans	NaN	Feuds, flirtations and toilet talk go down amo...	NaN	September 24, 2021	2021	TV-MA	1 Seaso		
8	s9	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train I...	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Season		
9	s10	International TV Shows, Reality TV	Feuds, flirtations and toilet talk go down amo...	NaN	Na	Na	Na	Na	Na	Na	Na	

In [249...]

```
# print(df.head())
df['country'] = df['country'].dropna().apply(lambda x : [i.strip() for i in x.split(',')])
df['listed_in'] = df['listed_in'].dropna().apply(lambda x : [i.strip() for i in x.split(',')])
df_genre = df.explode('listed_in')
df_genre = df_genre.explode('country')
top5_country = df_genre['country'].value_counts().index[:5]
top5_gen = df_genre['listed_in'].value_counts().index[:5]
top5_data = df_genre.loc[(df_genre["country"].isin(top5_country)) & (df_genre["listed_in"].isin(top5_gen))]
# print(top5_data)
top5_data = top5_data[['country', 'listed_in']].dropna().groupby(['country']).value_counts().reset_index()
# top5_data = top5_data.sort_values(ascending=False)
print(top5_data)
plt.figure(figsize=(15,10))
sns.barplot(x='country', y='count', hue='listed_in', data=top5_data)
plt.ylabel('Movie/Shows count')
plt.show()
```

	country	listed_in	count
0	Canada	Comedies	94
1	Canada	Dramas	82
2	Canada	International Movies	60
3	Canada	Action & Adventure	44
4	Canada	International TV Shows	25
5	France	International Movies	207
6	France	Dramas	167
7	France	Comedies	51
8	France	International TV Shows	43
9	France	Action & Adventure	37
10	India	International Movies	864
11	India	Dramas	662
12	India	Comedies	323
13	India	Action & Adventure	137
14	India	International TV Shows	66
15	United Kingdom	Dramas	197
16	United Kingdom	International Movies	170
17	United Kingdom	International TV Shows	128
18	United Kingdom	Comedies	91
19	United Kingdom	Action & Adventure	84
20	United States	Dramas	835
21	United States	Comedies	680
22	United States	Action & Adventure	404
23	United States	International Movies	166
24	United States	International TV Shows	74



Here we can see many insights about country and the genre they produce.

we can see that highest number of international movies are produced by India and higest number of drama is produced by USA.

BUSINESS INSIGHTS : netflix should focus more on drama related movies as they have higher demand then other genre.

#####

IV. On which day most of the movies/TV show are released.

```
In [251... pd.to_datetime(df['date_added'].str.strip(), dayfirst=True)
```

```
Out[251... 0    2021-09-25  
1    2021-09-24  
2    2021-09-24  
3    2021-09-24  
4    2021-09-24  
     ...  
8802   2019-11-20  
8803   2019-07-01  
8804   2019-11-01  
8805   2020-01-11  
8806   2019-03-02  
Name: date_added, Length: 8807, dtype: datetime64[ns]
```

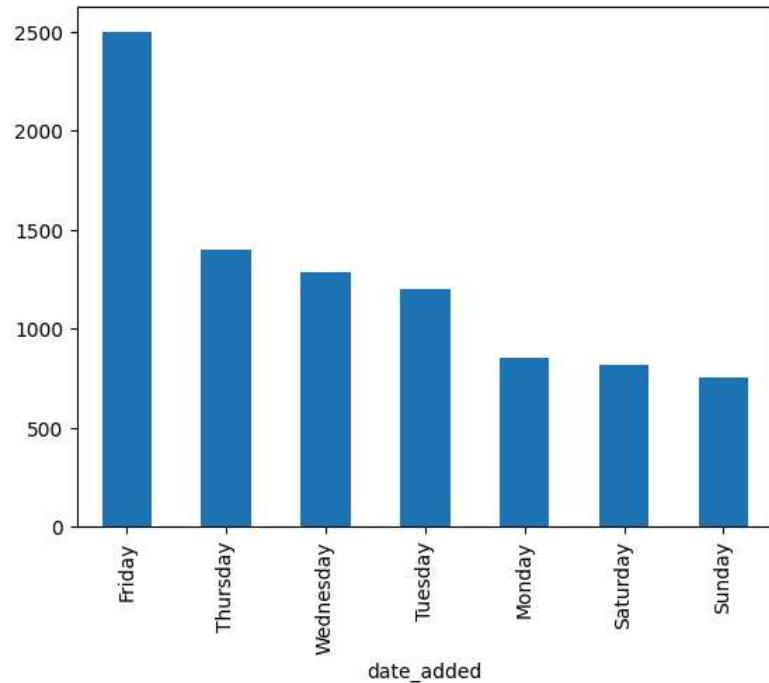
```
In [252... pd.to_datetime(df['date_added'].str.strip())
```

```
temp2 = pd.to_datetime(df['date_added'].str.strip())  
temp2.dt.weekday.value_counts()
```

```
Out[252... date_added  
4.0    2498  
3.0    1396  
2.0    1288  
1.0    1197  
0.0     851  
5.0     816  
6.0     751  
Name: count, dtype: int64
```

```
In [253... temp2.dt.day_name().value_counts().plot(kind='bar')
```

```
Out[253... <Axes: xlabel='date_added'>
```



Here we can clearly observe that most of the content is released in friday.

```
#####
V. Analysis of actors/directors of different types of shows/movies.
```

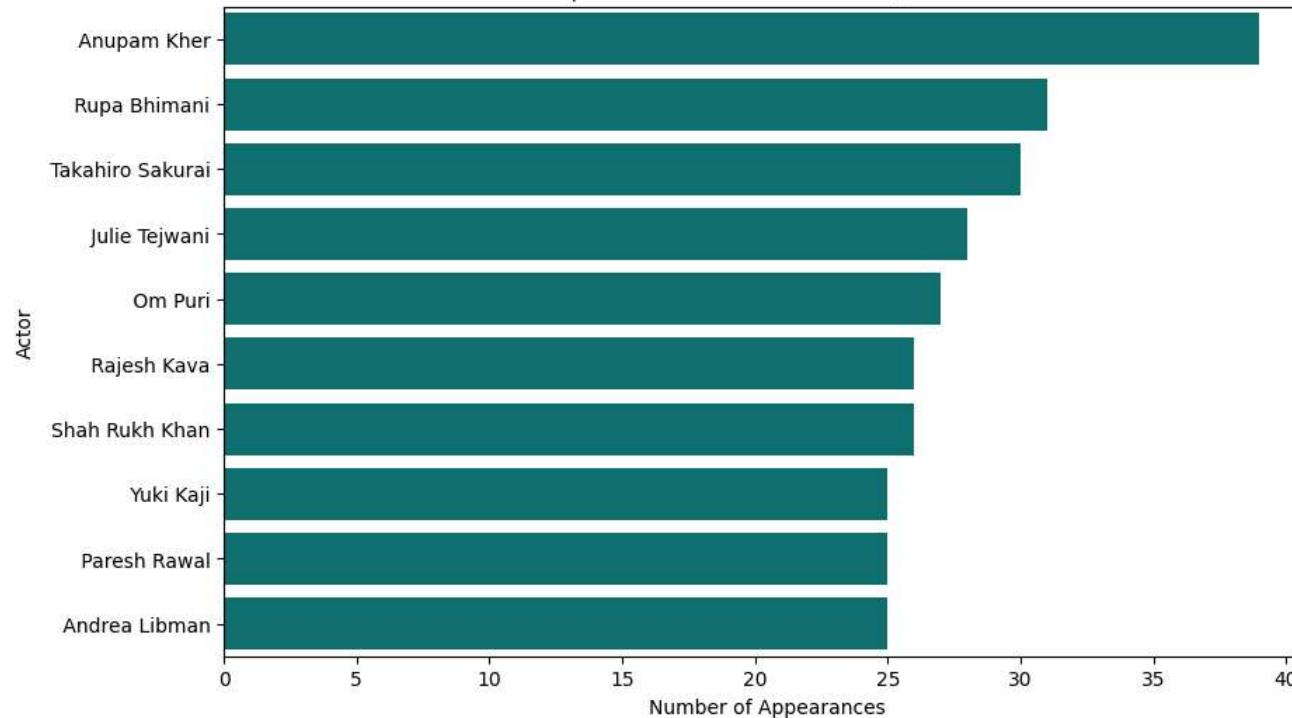
```
In [254]: top_actors = df['cast'].str.split(',').explode().value_counts().head(10)

top_directors = df['director'].value_counts().head(10)

# Plot the top 10 actors
plt.figure(figsize=(10, 6))
sns.barplot(x=top_actors.values, y=top_actors.index, color='teal')
plt.title('Top 10 Actors in Netflix Shows/Movies')
plt.xlabel('Number of Appearances')
plt.ylabel('Actor')
plt.show()

# Plot the top 10 directors
plt.figure(figsize=(10, 6))
sns.barplot(x=top_directors.values, y=top_directors.index, color='orange')
plt.title('Top 10 Directors in Netflix Shows/Movies')
plt.xlabel('Number of Shows/Movies Directed')
plt.ylabel('Director')
plt.show()
```

Top 10 Actors in Netflix Shows/Movies



Top 10 Directors in Netflix Shows/Movies

