```
import pandas as pd
import numpy as np
import math
import matplotlib.pyplot as plt
import seaborn as sns


url = "https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/000/940/original/netflix.csv"

netflix = pd.read_csv(url)


netflix.columns
```

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

```
netflix.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
```

```
netflix.shape
```

```
(8807, 12)
```

```
netflix.isna().sum()
```

```
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
```

```
netflix['duration']
```

```
0         125 min
1       9 Seasons
2         104 min
3         127 min
4         166 min
           ...
5327       96 min
5328      158 min
5329       88 min
5330       88 min
5331      111 min
Name: duration, Length: 5332, dtype: object
```

Dealing with Null Values

```
netflix = netflix.dropna().reset_index(drop=True)
netflix.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	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 photo shoot in Ghana, an American model s...
1	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 talented batch of amateur bakers face off in...
2	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 woman adjusting to life after a loss contend...
3	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 of her family is murdered in a terr...
4	s25	Movie	Jeans	S. Shankar	Prashanth, Aishwarya Rai	India	September 21,	1998	TV-14	166 min	Comedies, International	When the father of the man

```
data_du = netflix.copy()
data_du.groupby(['type'])['duration'].aggregate(['max', 'min'])
```

	max	min
type		
Movie	99 min	100 min
TV Show	9 Seasons	1 Season

```
netflix.shape
```

```
(5332, 12)
```

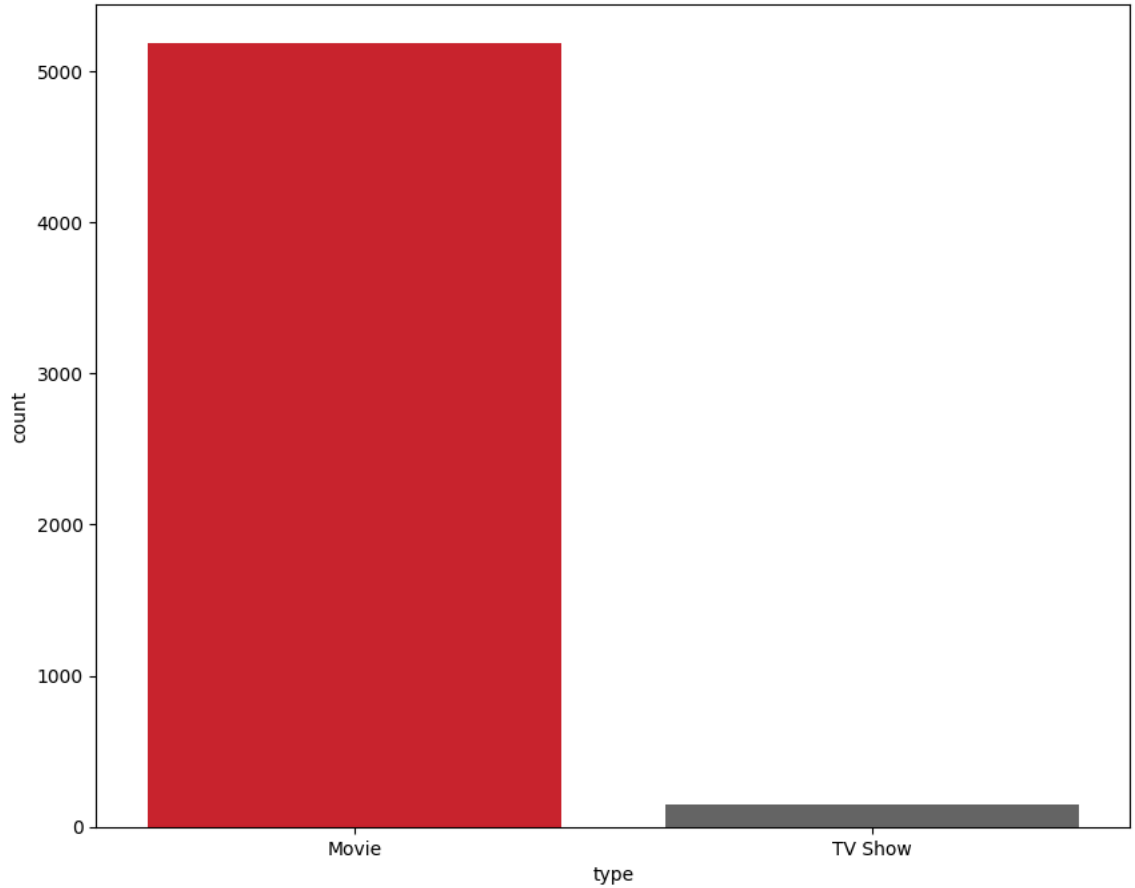
1. Find the counts of each categorical variable both using graphical and nongraphical analysis. Focusing on following categorical variables.
   a. type counts b. director counts c.country counts d.rating counts e. listed_in counts

```
#type counts
type_counts = netflix['type'].value_counts()
print(type_counts)
```

```
Movie      5185
TV Show     147
Name: type, dtype: int64
```

```
netflix_palette = ["#E50914", "#666666"]
plt.figure(figsize = (10,8))
sns.countplot(data = netflix, x='type',hue='type', palette=netflix_palette, legend =False)
```

```
<Axes: xlabel='type', ylabel='count'>
```



```
#b.director counts
director_counts = netflix['director'].value_counts()
print(director_counts)
```

```
     Raúl Campos, Jan Suter           18
     Marcus Raboy                     15
     Jay Karas                        14
     Cathy Garcia-Molina              13
     Martin Scorsese                  12
                                      ..
     Igor Kovalyov, Norton Virgien     1
     Danny Cannon                      1
     Mana Yasuda                       1
     Gupse Özay                        1
     Mozez Singh                       1
     Name: director, Length: 3945, dtype: int64
```

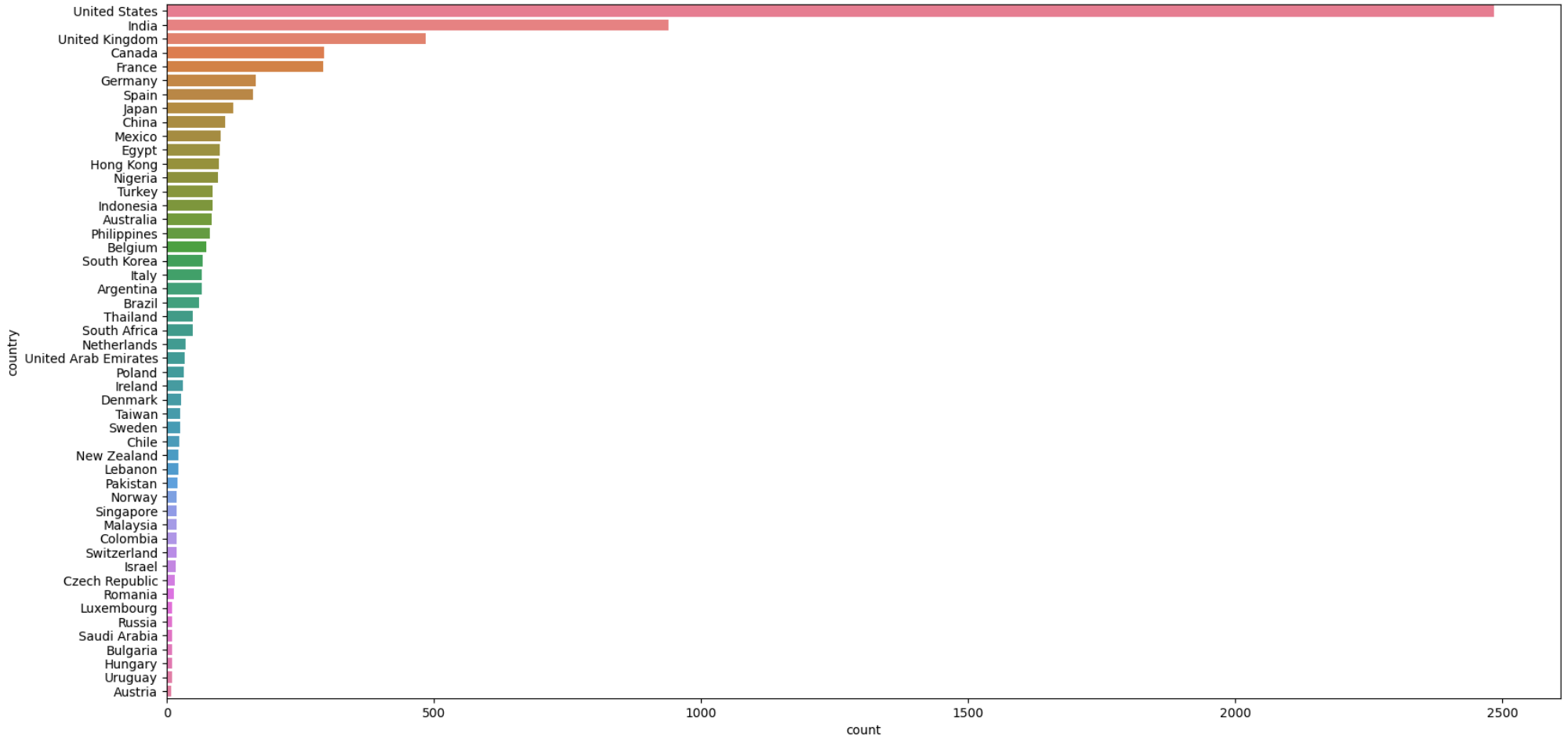
```
#c.Country wise contents
data_country = netflix
data_country['country'] = data_country['country'].str.split(',')
data_country = data_country.explode('country')
data_country['country'] = data_country['country'].str.strip()

country_count = data_country.groupby(['country'])['title'].nunique().reset_index(name = "count")



#having 175 countries, taking only top 50 for better readability
top_50countries = country_count.sort_values(by='count', ascending = False).head(50)

top_50countries
```

	country	count
103	United States	2485
40	India	940
102	United Kingdom	485
17	Canada	295
31	France	293
33	Germany	167
92	Spain	161
48	Japan	124
20	China	109
61	Mexico	101
28	Egypt	99
37	Hong Kong	98
70	Nigeria	96
100	Turkey	86
41	Indonesia	85
6	Australia	84
76	Philippines	81
10	Belgium	74
90	South Korea	68
46	Italy	65
5	Argentina	65
12	Brazil	60
99	Thailand	48
89	South Africa	48
67	Netherlands	35
101	United Arab Emirates	34
77	Poland	32
44	Ireland	31
24	Denmark	27
98	Taiwan	26
95	Sweden	25
19	Chile	24
68	New Zealand	22

53	Lebanon	22
72	Pakistan	20
71	Norway	19
85	Singapore	19
58	Malaysia	19
21	Colombia	19
96	Switzerland	18
45	Israel	17
23	Czech Republic	15
80	Romania	13
56	Luxembourg	10
81	Russia	10
82	Saudi Arabia	10
13	Bulgaria	10
38	Hungary	10
104	Uruguay	10
7	Austria	9

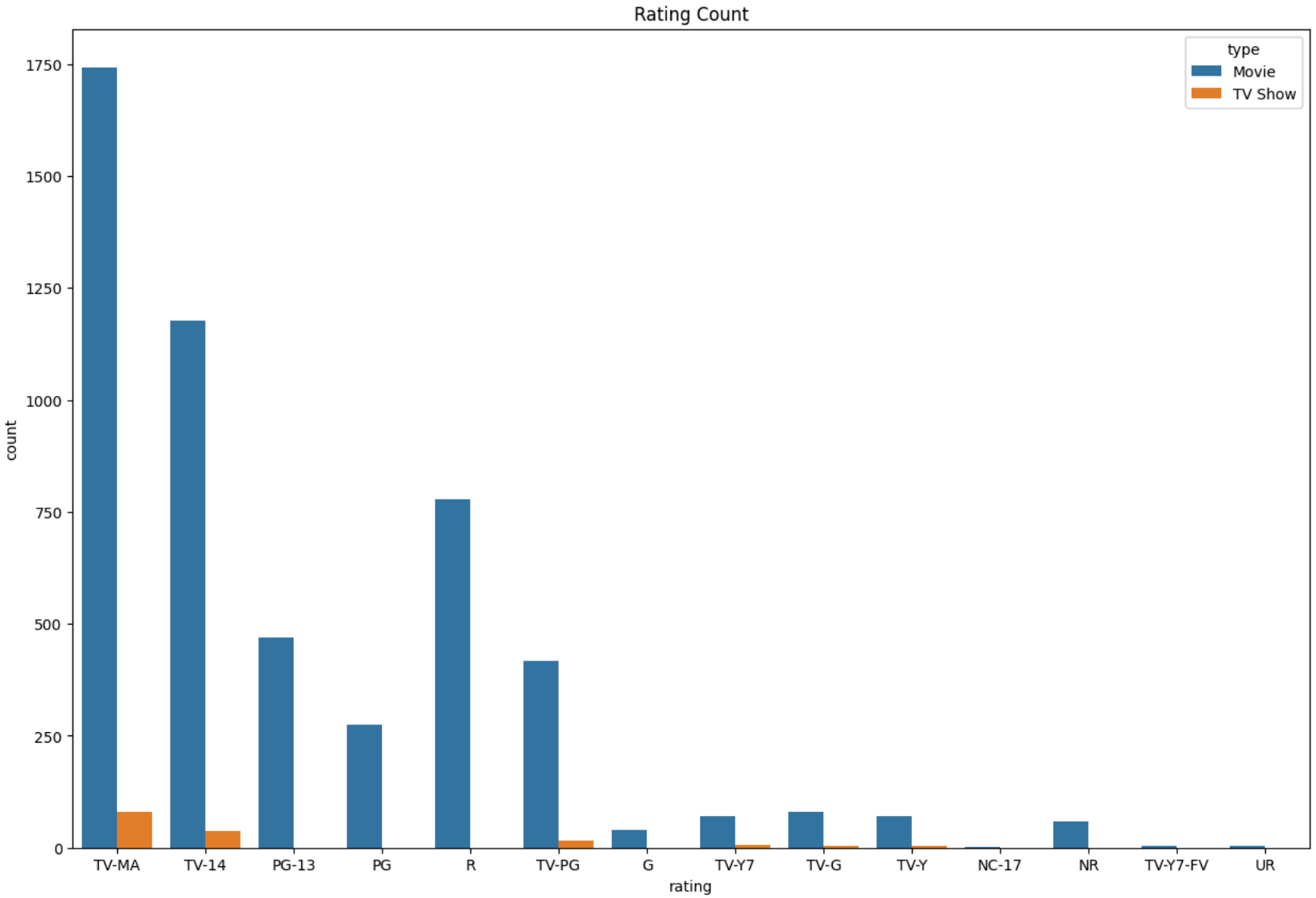
```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize = (20,10))
sns.barplot(data=top_50countries, y = 'country', x='count', hue='country')
```

```
<Axes: xlabel='count', ylabel='country'>
```




```
plt.figure(figsize = (15,10))
sns.countplot(data=netflix, x='rating', hue ='type')
plt.title('Rating Count')
```

```
Text(0.5, 1.0, 'Rating Count')
```




```
#Genre counting
data_genre = netflix.copy()
```

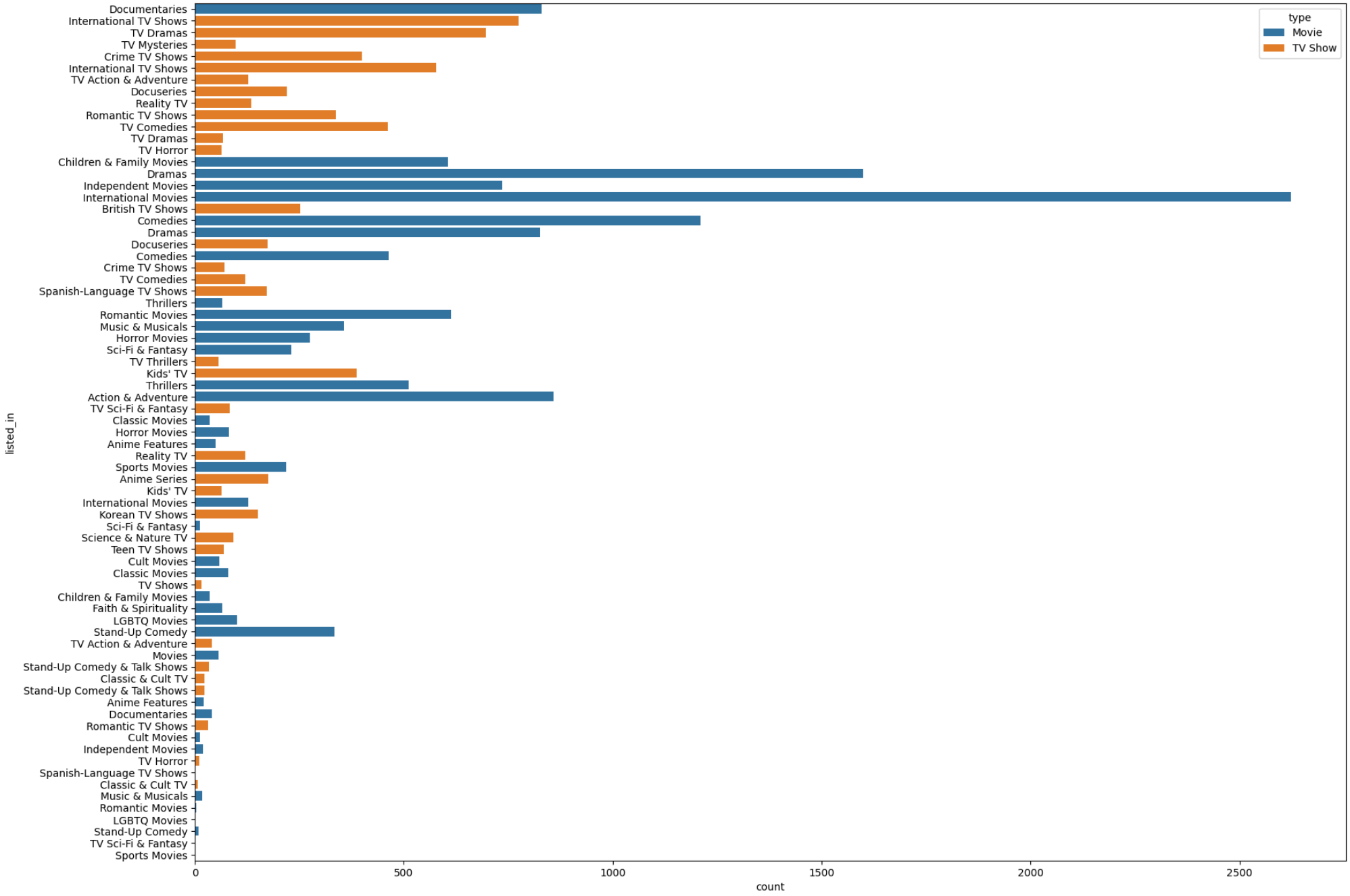
```
data_genre['listed_in'] = data_genre['listed_in'].str.split(',')
data_genre = data_genre.explode('listed_in')
genre_count = data_genre.groupby(['listed_in'])['title'].nunique().reset_index(name = "count")
genre_count
```

	listed_in	count
0	Anime Features	43
1	Children & Family Movies	34
2	Classic & Cult TV	2
3	Classic Movies	35
4	Comedies	426
...	...	...
65	TV Comedies	3
66	TV Dramas	1
67	TV Horror	1
68	TV Shows	5
69	Thrillers	62

70 rows × 2 columns

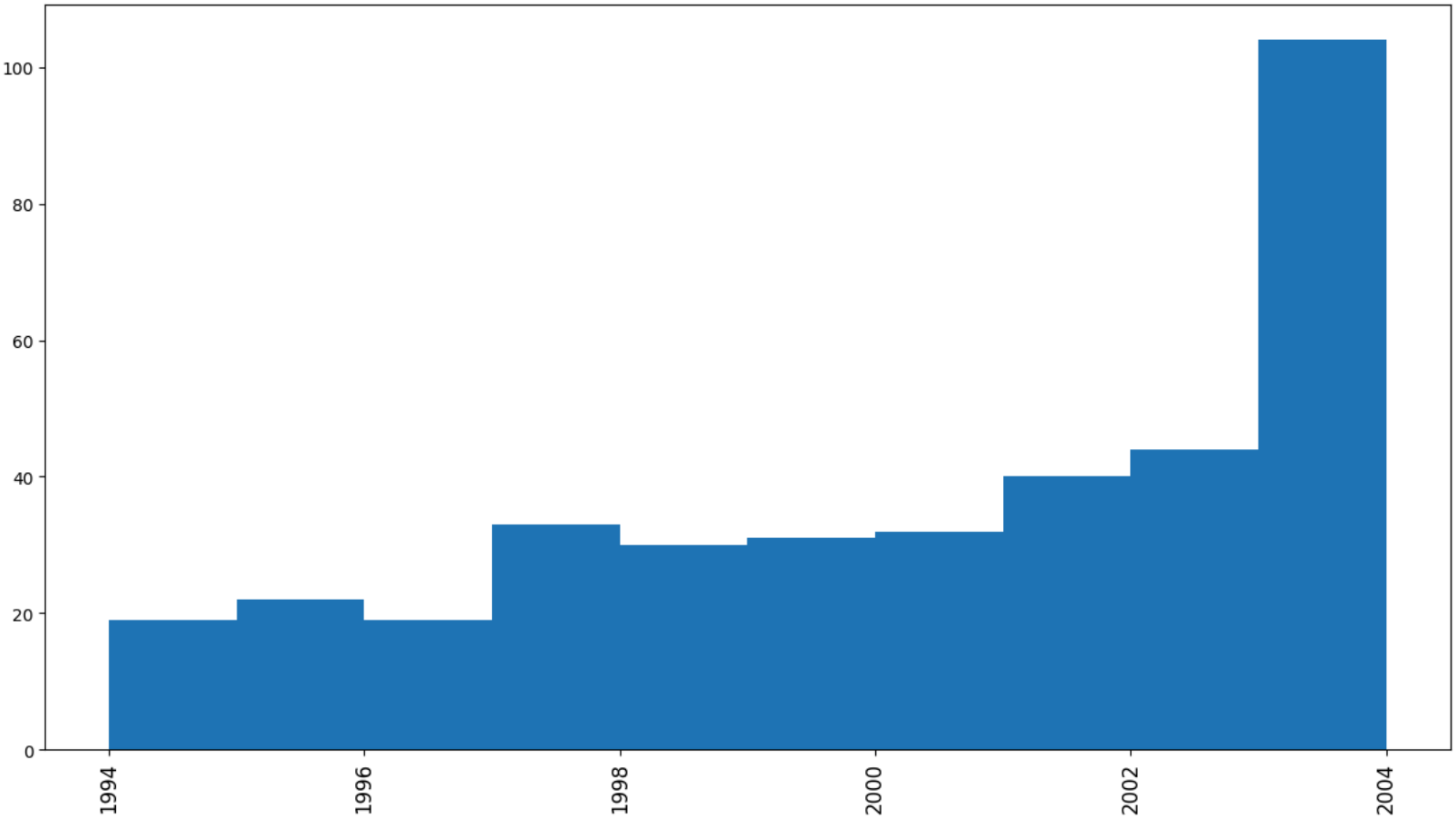
```
plt.figure(figsize = (20,15))
sns.countplot(data=data_genre, y='listed_in',hue='type')
```

```
<Axes: xlabel='count', ylabel='listed_in'>
```



```
# 1. How has the number of movies released per year changed over the last 20-30 years?
data_movies = netflix.loc[netflix['type']=='Movie']
current_year = pd.Timestamp.now().year
last_20yr_data = netflix[(netflix['release_year']>= current_year -30) & (netflix['release_year']<= current_year- 20)]
plt.figure(figsize = (15,8))
plt.hist(last_20yr_data['release_year'], bins=10)
plt.xticks(rotation=90, fontsize=12)
```

```
(array([1992., 1994., 1996., 1998., 2000., 2002., 2004., 2006.]),
 [Text(1992.0, 0, '1992'),
  Text(1994.0, 0, '1994'),
  Text(1996.0, 0, '1996'),
  Text(1998.0, 0, '1998'),
  Text(2000.0, 0, '2000'),
  Text(2002.0, 0, '2002'),
  Text(2004.0, 0, '2004'),
  Text(2006.0, 0, '2006')])
```



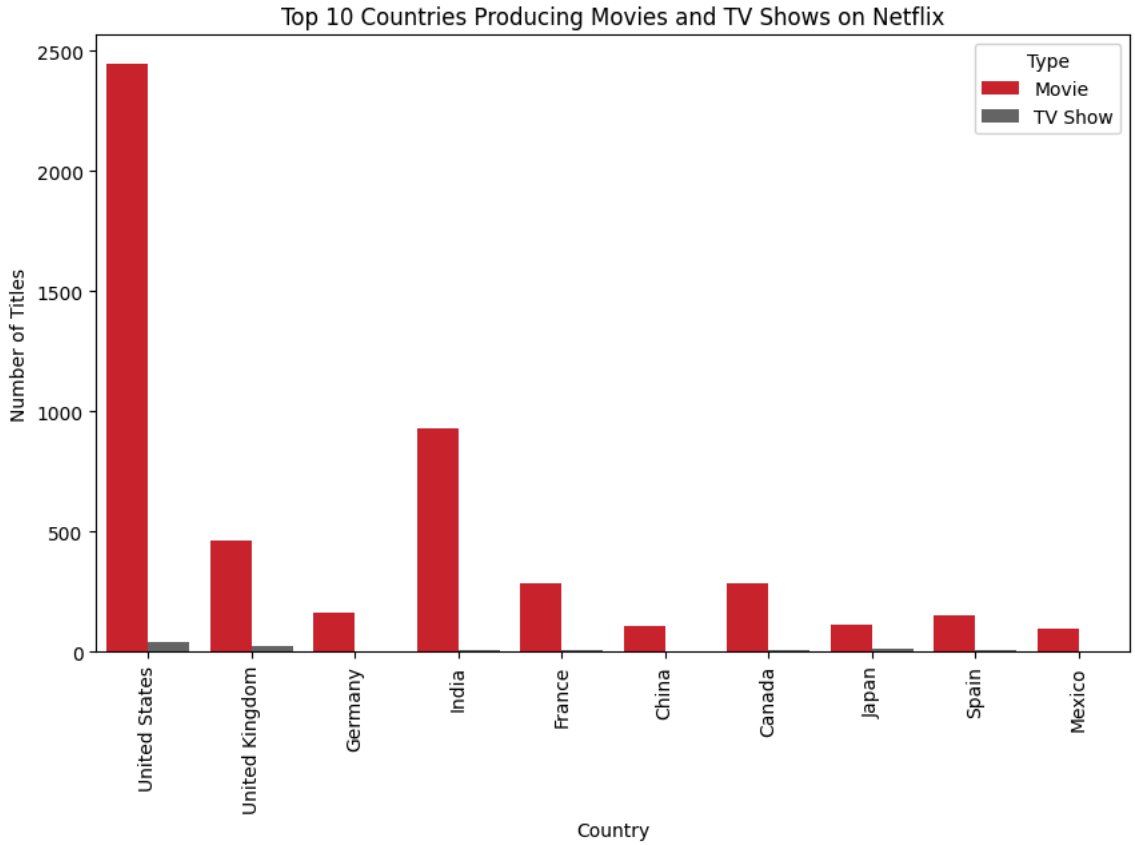
2. Comparison of tv shows vs. movies.

```
x = country_count.sort_values(by="count", ascending=False).head(10)
data_country_top_10 = data_country[data_country['country'].isin(x.country)]
data_country_top_10
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States	September 24, 2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s...
0	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United Kingdom	September 24, 2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s...
0	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	Germany	September 24, 2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s...
1	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 talented batch of amateur bakers face off in...
2	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 woman adjusting to life after a loss contend...
...	...	...	...	...	...	...	...	...	...	...	...	...
5326	s8800	Movie	Zenda	Avadhoot Gupte	Santosh Juvekar, Siddharth Chandekar, Sachit P...	India	February 15, 2018	2009	TV-14	120 min	Dramas, International Movies	A change in the leadership of a political part...
5328	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...
5329	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...
5330	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
5331	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...

5160 rows × 12 columns

```
plt.figure(figsize=(10, 6))
sns.countplot(x='country', hue='type', data=data_country_top_10, palette=netflix_palette)
plt.title('Top 10 Countries Producing Movies and TV Shows on Netflix')
plt.xlabel('Country')
plt.ylabel('Number of Titles')
plt.xticks(rotation=90)
plt.legend(title='Type')
plt.show()
```



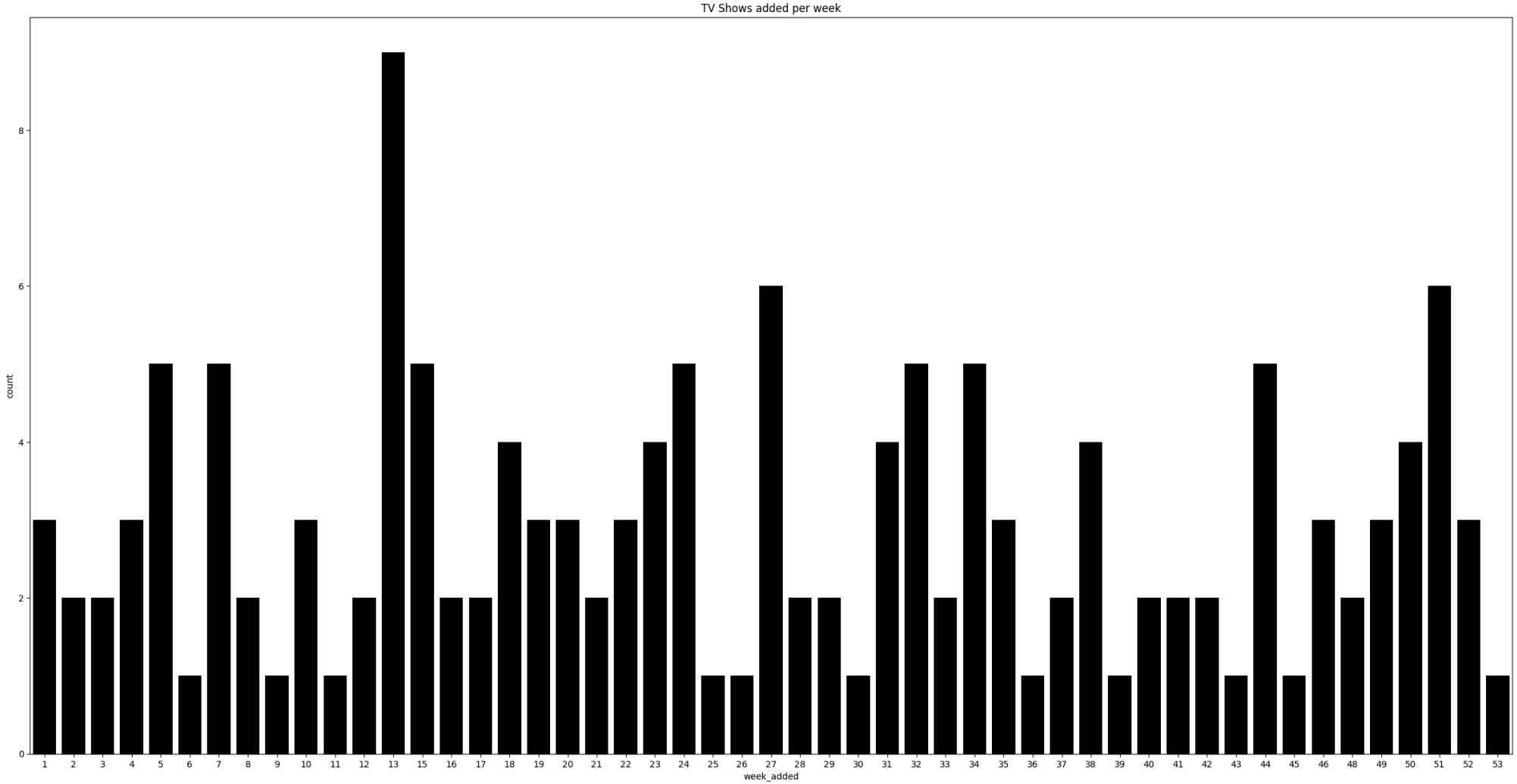
3.What is the best time to launch a TV show?

```
data = netflix.copy()
data['date_added'] = pd.to_datetime(data['date_added'])
data['year_added'] = data['date_added'].dt.strftime('%Y').astype('int64')
data['month_added'] = data['date_added'].dt.strftime('%b')
data['day_added'] = data['date_added'].dt.strftime('%d').astype('int64')
data['dayname_added'] = data['date_added'].dt.strftime('%a')
data['week_added'] = data['date_added'].dt.isocalendar().week
```

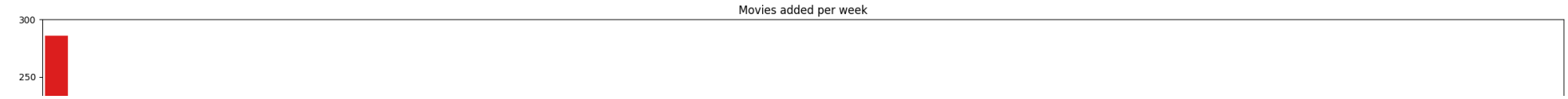
```
data_movies = data.loc[data['type']=='Movie']
data_TV = data.loc[data['type']=='TV Show']
#For TV Shows
tv_weekly_count = data_TV.groupby(['week_added'])['title'].nunique().reset_index(name='count').sort_values(by='count')
tv_monthly_count = data_TV.groupby(['month_added'])['title'].nunique().reset_index(name='count').sort_values(by='count')

#for Movies
movie_weekly_count = data_movies.groupby(['week_added'])['title'].nunique().reset_index(name='count').sort_values(by='count')
movie_monthly_count = data_movies.groupby(['month_added'])['title'].nunique().reset_index(name='count').sort_values(by='count')



plt.figure(figsize=(20,15)).suptitle("Best time to Launch content",fontsize=20)

plt.figure(figsize=(30,15)).suptitle("Best time to Launch content",fontsize=20)
plt.subplot(1,1,1)
sns.barplot(data = tv_weekly_count, x = 'week_added', y='count', color = 'black')
plt.title('TV Shows added per week')

plt.figure(figsize=(30,15)).suptitle("Best time to Launch content",fontsize=20)
plt.subplot(2,1,1)
sns.barplot(data = movie_weekly_count, x = 'week_added', y='count', color = 'red')
plt.title('Movies added per week')

plt.subplot(2,3,4)
sns.barplot(data = tv_monthly_count, x = 'month_added', y='count', color = 'black')
plt.title('TV Shows added per month')


plt.subplot(2,3,6)
sns.barplot(data = movie_monthly_count, x = 'month_added', y='count', color = 'red')
plt.title('Movies added per month')
```

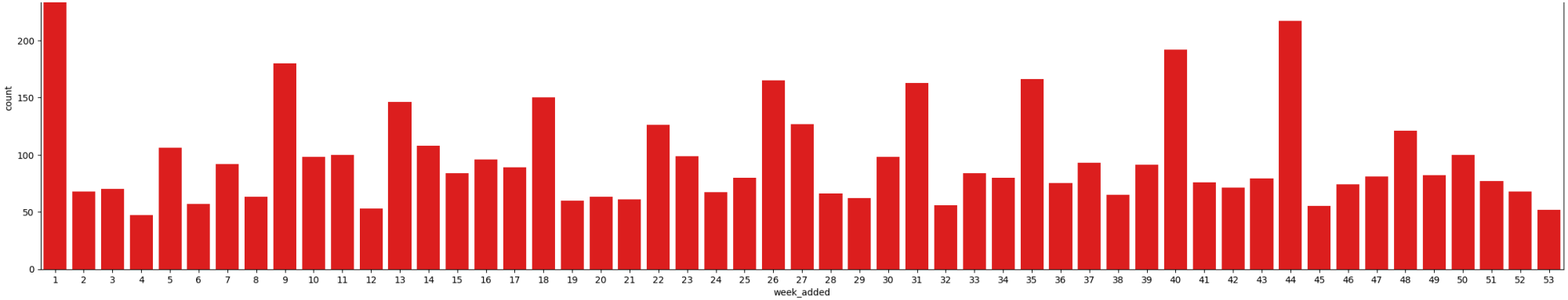
```
Text(0.5, 1.0, 'Movies added per month')
<Figure size 2000x1500 with 0 Axes>
```

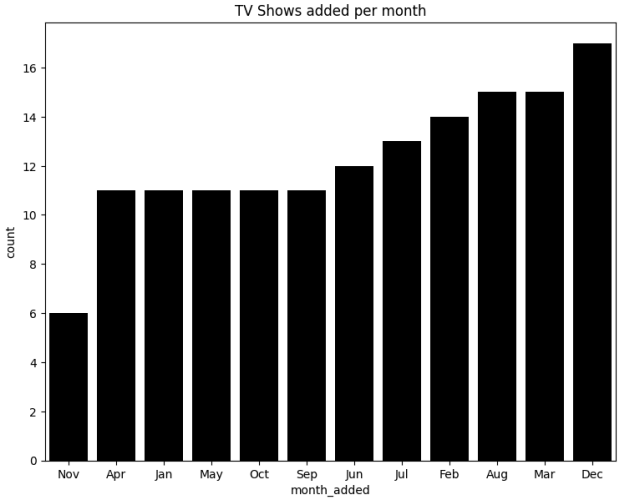
Best time to Launch content

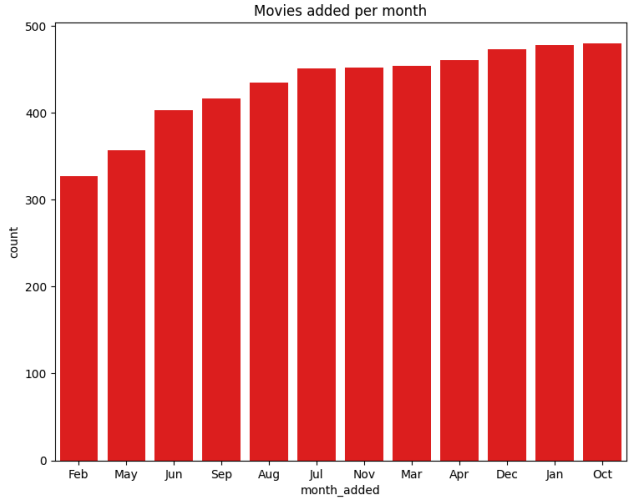


Best time to Launch content




count
week_added


TV Shows added per month
count
month_added
Nov
Apr
Jan
May
Oct
Sep
Jun
Jul
Feb
Aug
Mar
Dec


Movies added per month
count
month_added
Feb
May
Jun
Sep
Aug
Jul
Nov
Mar
Apr
Dec
Jan
Oct

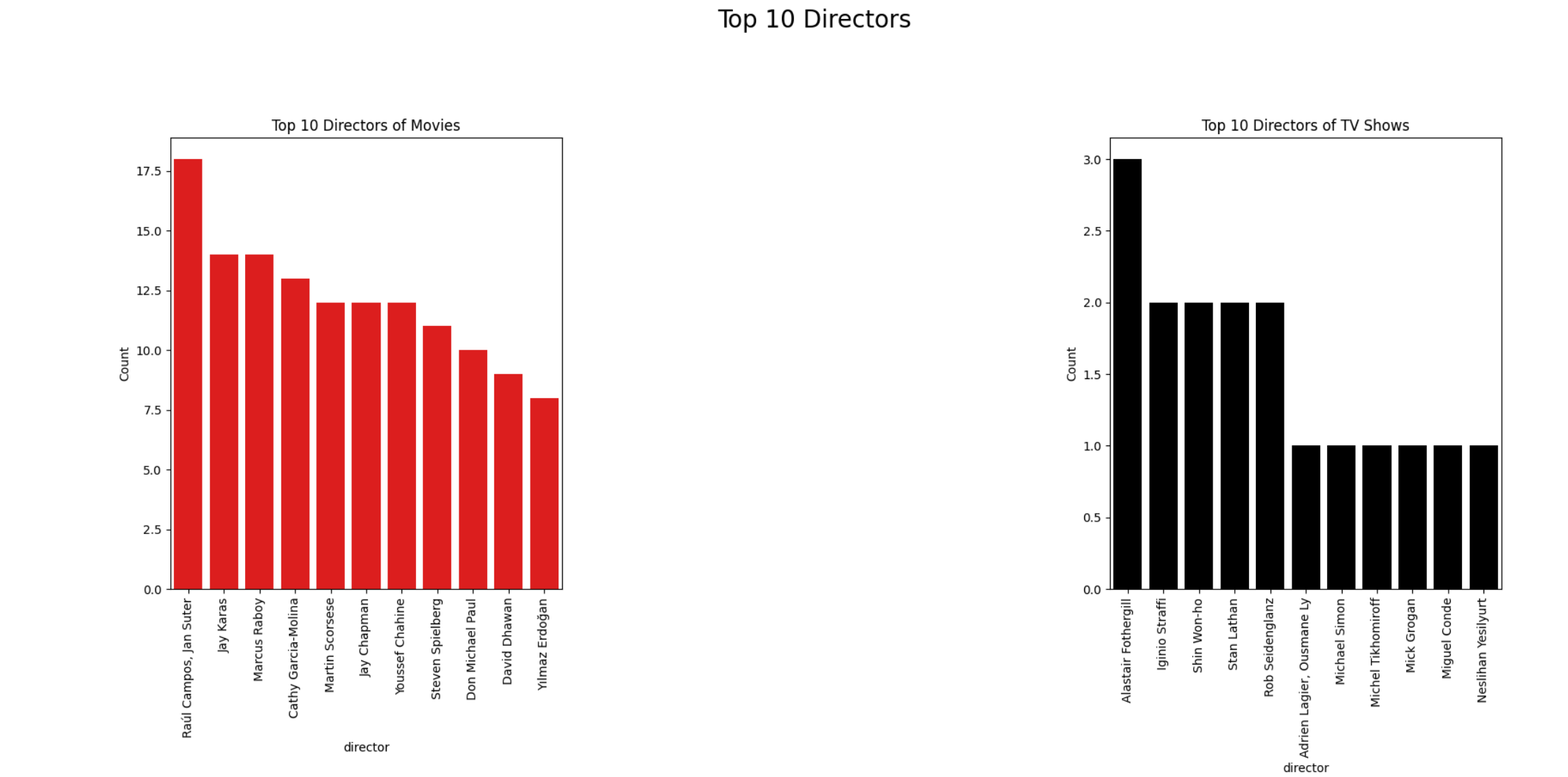
4. Analysis of actors/directors of different types of shows/movies.

```
# a. Identify the top 10 directors who have appeared in most movies or TV shows.
# For Movies
directors_mv = data_movies.groupby(['director'])['title'].nunique().sort_values(ascending =False)
Top_movie_directors = directors_mv.head(11).reset_index(name ='Count')
print(Top_movie_directors)

# For TV_shows
directors_tv = data_TV.groupby(['director'])['title'].nunique().sort_values(ascending = False)
Top_TV_directors = directors_tv.head(11).reset_index(name = 'Count')
print(Top_TV_directors )
```

```
                  director  Count
0   Raúl Campos, Jan Suter     18
1                Jay Karas     14
2             Marcus Raboy     14
3      Cathy Garcia-Molina     13
4          Martin Scorsese     12
5              Jay Chapman     12
6           Youssef Chahine     12
7         Steven Spielberg     11
8         Don Michael Paul     10
9             David Dhawan      9
10          Yılmaz Erdoğan      8
                     director  Count
0         Alastair Fothergill      3
1              Iginio Straffi      2
2                 Shin Won-ho      2
3                 Stan Lathan      2
4              Rob Seidenglanz      2
5   Adrien Lagier, Ousmane Ly      1
6                Michael Simon      1
7          Michel Tikhomiroff      1
8                 Mick Grogan      1
9                Miguel Conde      1
10         Neslihan Yesilyurt      1
```

```
plt.figure(figsize=(20,15)).suptitle("Top 10 Directors",fontsize=20)

plt.subplot(2,3,1)
sns.barplot(data = Top_movie_directors, x = 'director', y='Count', color = 'red')
plt.title('Top 10 Directors of Movies')
plt.xticks(rotation = 90)


plt.subplot(2,3,3)
sns.barplot(data = Top_TV_directors, x = 'director', y='Count', color ='black')
plt.title('Top 10 Directors of TV Shows')
plt.xticks(rotation = 90)
```

```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
 [Text(0, 0, 'Alastair Fothergill'),
  Text(1, 0, 'Iginio Straffi'),
  Text(2, 0, 'Shin Won-ho'),
  Text(3, 0, 'Stan Lathan'),
  Text(4, 0, 'Rob Seidenglanz'),
  Text(5, 0, 'Adrien Lagier, Ousmane Ly'),
  Text(6, 0, 'Michael Simon'),
  Text(7, 0, 'Michel Tikhomiroff'),
  Text(8, 0, 'Mick Grogan'),
  Text(9, 0, 'Miguel Conde'),
  Text(10, 0, 'Neslihan Yesilyurt')])
```



```
# b. Identify the top 10 actors who have appeared in most movies or TV shows.
# For Movies
data_cast_mv = data_movies.copy()
data_cast_mv['cast'] = data_cast_mv['cast'].str.split(',')
data_cast_mv = data_cast_mv.explode('cast')
actors_mv = data_cast_mv.groupby(['cast'])['title'].nunique().sort_values(ascending =False)
Top_movie_actors = actors_mv.head(10).reset_index(name ='Count')



# For TV Shows
data_cast_tv = data_TV.copy()
data_cast_tv['cast'] = data_cast_tv['cast'].str.split(',')
data_cast_tv = data_cast_tv.explode('cast')
actors_tv = data_cast_tv.groupby(['cast'])['title'].nunique().sort_values(ascending = False)
Top_TV_actors = actors_tv.head(10).reset_index(name = 'Count')



plt.figure(figsize=(20,15)).suptitle("Top 10 Actors",fontsize=20)

plt.subplot(2,3,1)
sns.barplot(data = Top_movie_actors, x = 'cast', y='Count', color = "red")
plt.title('Top 10 Actors of Movies')
plt.xticks(rotation = 90)


plt.subplot(2,3,3)
sns.barplot(data = Top_TV_actors, x = 'cast', y='Count', color = "black")
plt.title('Top 10 Actors of TV Shows')
plt.xticks(rotation = 90)
```