

```
In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [3]: # Reading the data set
df = pd.read_csv(r'C:\Users\ssd\OneDrive\Documents\Numpy notes\11_Pandas_3\11_Pandas_3\netflix.csv')
```

```
In [4]: # splitting the fields and exploding them i.e stacking them so that the data gets uniformly arranged in the data frame
df['director'] =df['director'].str.split(',')
```

```
In [5]: df = df.explode(['director'])
```

```
In [6]: df['cast'] =df['cast'].str.split(',')
```

```
In [7]: df = df.explode(['cast'])
```

```
In [8]: df['listed_in'] =df['listed_in'].str.split(',')
```

```
In [9]: df = df.explode(['listed_in'])
```

```
In [10]: # Analyzing the data set  
df
```

Out[10]:

	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	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	TV Mysteries	After crossing paths at a party, a Cape Town t...
1	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...
...	...	...	...	...	...	...	...	...	...	...	...	...
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Dramas	A scrappy but poor boy worms his way into a ty...
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...

161216 rows × 12 columns

```
In [11]: # we can see that date format is in object format and need to change it to data format
# Type casting will be done in the next step
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 161216 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype  
---  -
 0   show_id         161216 non-null object  
 1   type            161216 non-null object  
 2   title           161216 non-null object  
 3   director        116595 non-null object  
 4   cast            159483 non-null object  
 5   country         149319 non-null object  
 6   date_added      161058 non-null object  
 7   release_year    161216 non-null int64   
 8   rating          161149 non-null object  
 9   duration        161213 non-null object  
10   listed_in       161216 non-null object  
11   description      161216 non-null object  
dtypes: int64(1), object(11)
memory usage: 16.0+ MB
```

```
In [12]: df.describe()  
# checking for the numerical summary of data in the data set
```

```
Out[12]:
```

	release_year
count	161216.000000
mean	2013.649675
std	9.135247
min	1925.000000
25%	2012.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

```
In [13]: # checking for null values in the data frame
# regular isna gives the boolean vales representing False if no null present, True if null exists in that place
# but how do we get overall count of null values categorized by columns and rows ?
df.isna()
```

```
Out[13]:
```

	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
1	False	False	False	True	False	False	False	False	False	False	False	False
1	False	False	False	True	False	False	False	False	False	False	False	False
1	False	False	False	True	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...
8806	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
8806	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
8806	False	False	False	False	False	False	False	False	False	False	False	False

161216 rows × 12 columns

```
In [14]: # column wise null count  
# sum takes default axis value as 0  
df.isna().sum()
```

```
Out[14]: show_id          0  
type          0  
title         0  
director     44621  
cast         1733  
country     11897  
date_added   158  
release_year  0  
rating       67  
duration      3  
listed_in    0  
description   0  
dtype: int64
```

```
In [76]: # since we are good with gathering the null count of rows and columns, how do we handle them
# this is where cleaning and polishing phase of pandas begins, we clean and polish the null values now
# for that we use an important method in pandas for replacing the null values and that is fillna()

df['director'] = df['director'].fillna(df['director'].mode()[0])
df['cast'] = df['cast'].fillna(df['cast'].mode()[0])
df['country'] = df['country'].fillna(df['country'].mode()[0])
df['date_added'] = df['date_added'].fillna(df['date_added'].mode()[0])
df['rating'] = df['rating'].fillna(df['rating'].mode()[0])
df['duration'] = df['duration'].fillna(df['duration'].mode()[0])

#Converting data types appropriately
df["date_added"] = df["date_added"].astype('datetime64[ns]')

# for the further analysis on year month and month based on the content added in Netflix
df['year'] = df['date_added'].dt.strftime('%Y')
df['month'] = df['date_added'].dt.strftime('%m')

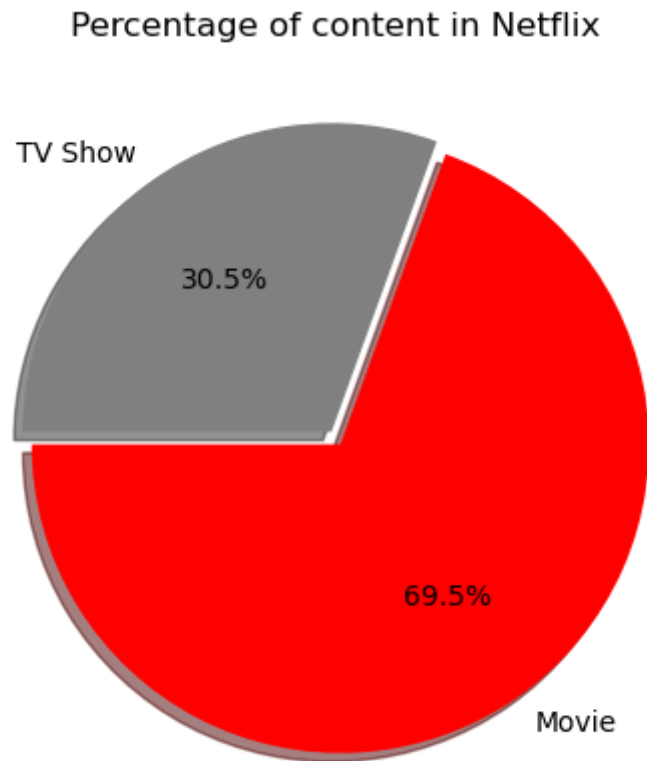
df['year'] = df['year'].astype('int')
df['month'] = df['month'].astype('int')
df.info()

# with that we have cleaned and polished our data
# Lets go on with ETA
```



```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 161216 entries, 0 to 8806
Data columns (total 14 columns):
#   Column          Non-Null Count  Dtype
---  -
0   show_id         161216 non-null object
1   type            161216 non-null object
2   title           161216 non-null object
3   director        161216 non-null object
4   cast            161216 non-null object
5   country         161216 non-null object
6   date_added      161216 non-null datetime64[ns]
7   release_year    161216 non-null int64
8   rating          161216 non-null object
9   duration        161216 non-null object
10  listed_in       161216 non-null object
11  description      161216 non-null object
12  year            161216 non-null int32
13  month           161216 non-null int32
dtypes: datetime64[ns](1), int32(2), int64(1), object(10)
memory usage: 17.2+ MB
```

```
In [18]: # Let us Look at content in Netflix categorized by movies and TV shows till date
# since the descriptive stats always better understood with Visualization, Lets understand it that way
plt.figure(figsize = (8,5))
plt.title('Percentage of content in Netflix')
plt.pie(df['type'].value_counts(),
        explode = (0.025,0.025),labels = df['type'].value_counts().index,
        colors = ['red','grey'],autopct='%1.1f%%', startangle=180, shadow = True )
plt.show()
# From this particular visualization
# it is quite evident that the amount of movies are comparatively more than TV shows in Netflix
```



```
In [19]: year_wise_movies_count = df['release_year'].value_counts()  
year_wise_movies_count.index  
year_wise_movies_count.values  
year_wise_movies_count.index.max()
```

Out[19]: 2021

```
In [20]: # checking for the trend of movies that has been released from the past 30 years  
  
movies_released_in_last_30_years = df.loc[(df['release_year'] >= 1991) & (df['release_year'] <= 2021) &  
                                           (df['type'] == 'Movie')]
```

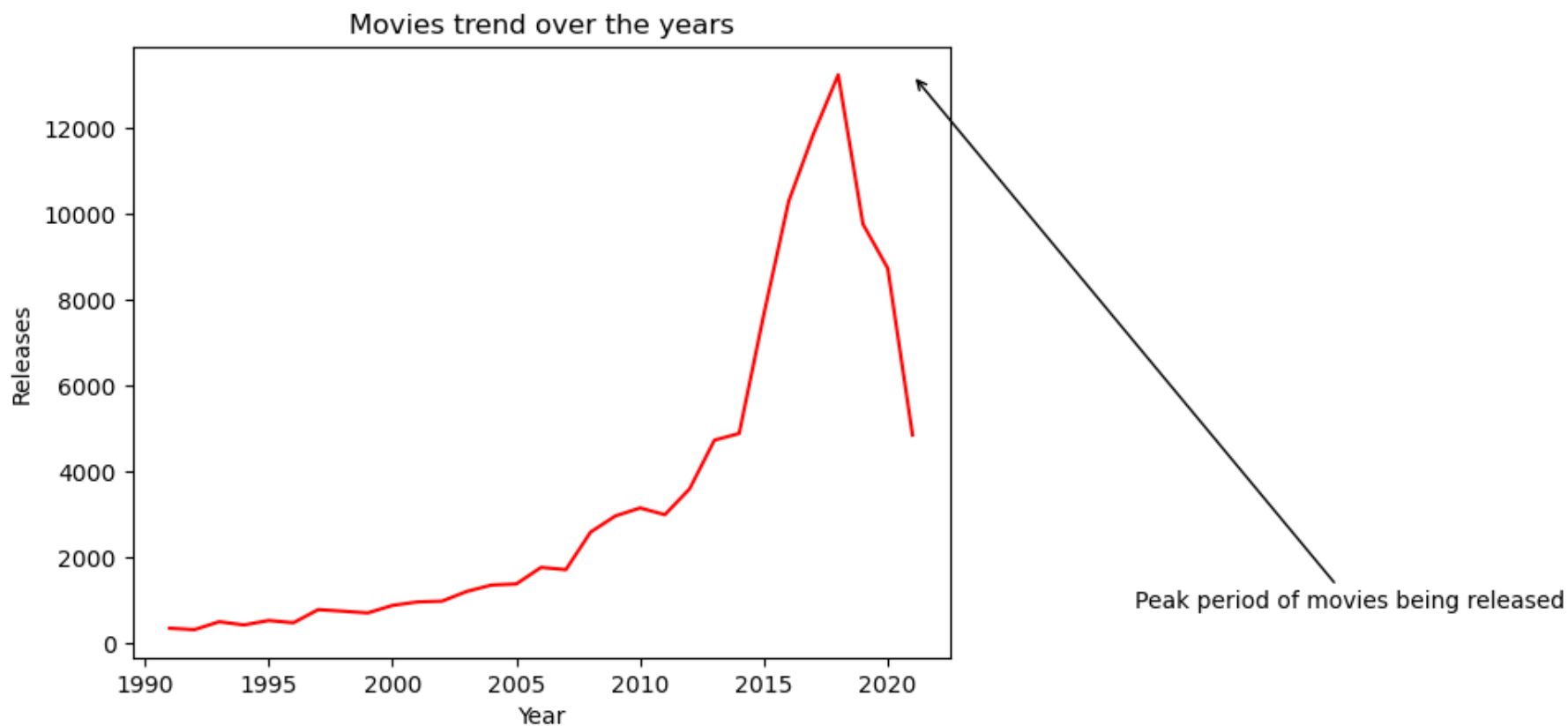
```
In [21]: date_between_1991_and_2021 = movies_released_in_last_30_years['release_year'].value_counts()
```

```
In [22]: date_between_1991_and_2021.index.max()
```

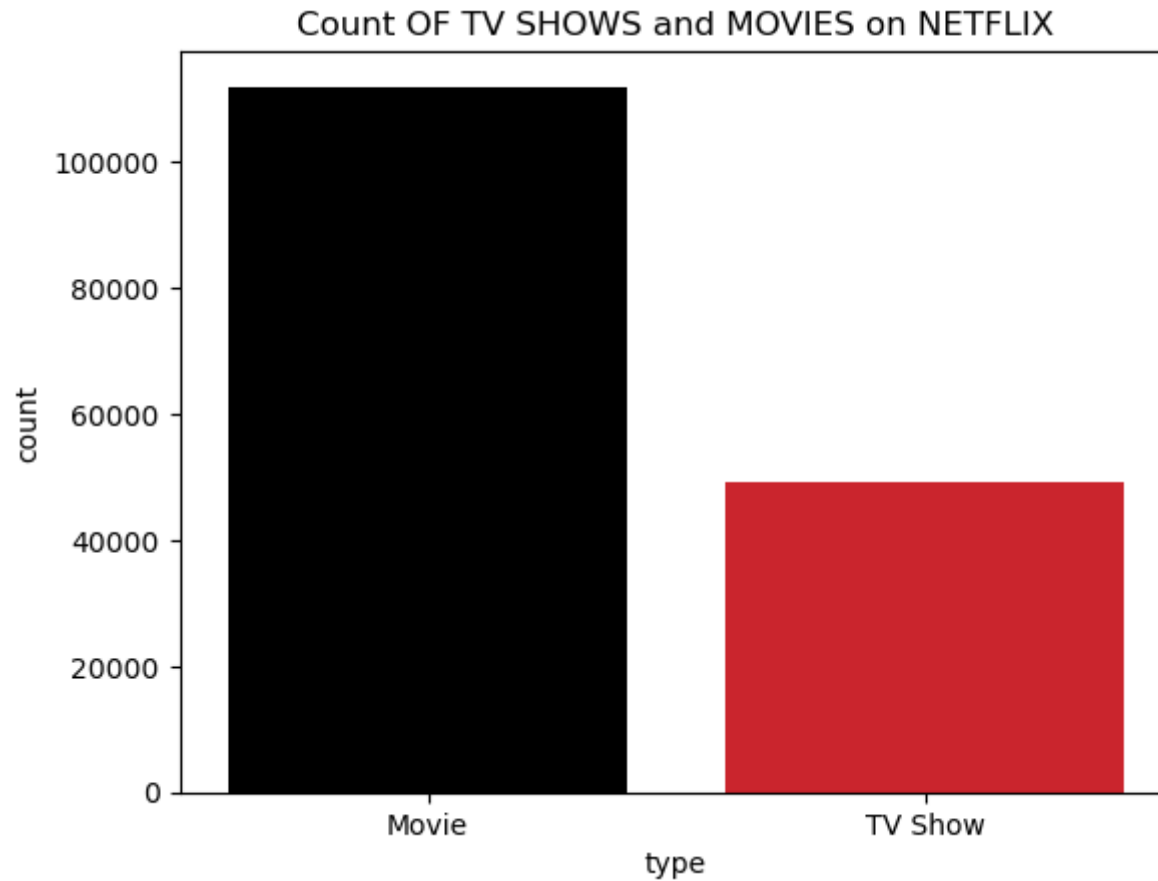
Out[22]: 2021

In [23]: *# Here we actually used two series as the x and y axis, so it is possible to use series of index and values in the line plot*  
*# x represents the year and y represents the counts*

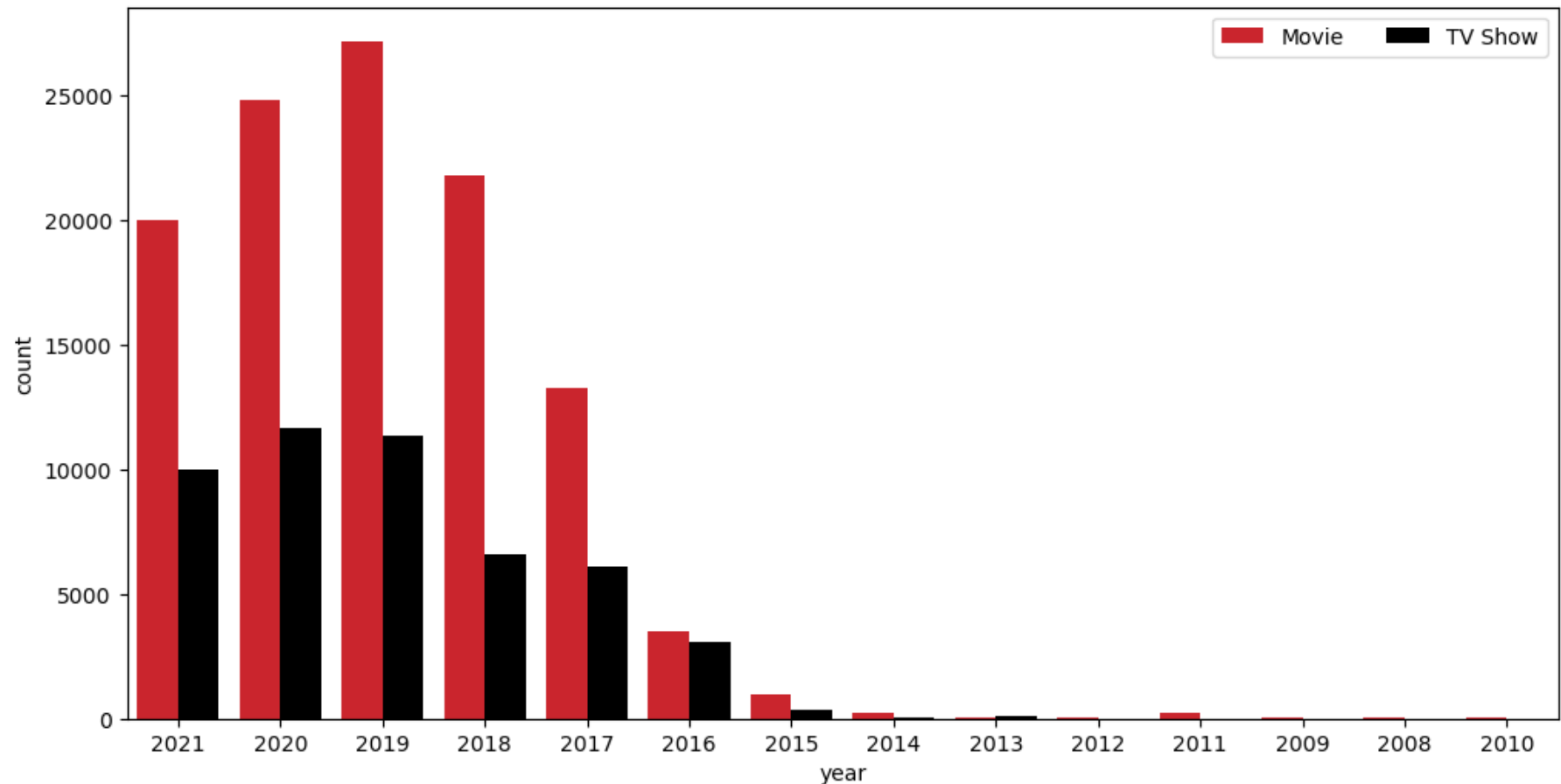
```
sns.lineplot(x = date_between_1991_and_2021.index, y = date_between_1991_and_2021.values,color = 'red')
plt.title('Movies trend over the years')
plt.annotate('Peak period of movies being released',
            xy = (date_between_1991_and_2021.index.max(),date_between_1991_and_2021.values.max()),
            xytext = (2030,800),
            arrowprops={"arrowstyle":"->"})
plt.ylabel('Releases')
plt.xlabel('Year')
plt.show()
# From the past 30 years there is significant increment in the number of movies being released till early 2019
```



```
In [24]: # Count of movies being streaming in the Netflix till date  
sns.countplot(data = df, x = 'type',palette=["#000000", '#E50914']).set(title='Count OF TV SHOWS and MOVIES on NETFLIX')  
plt.show()  
# count of movies are greater than tv shows
```

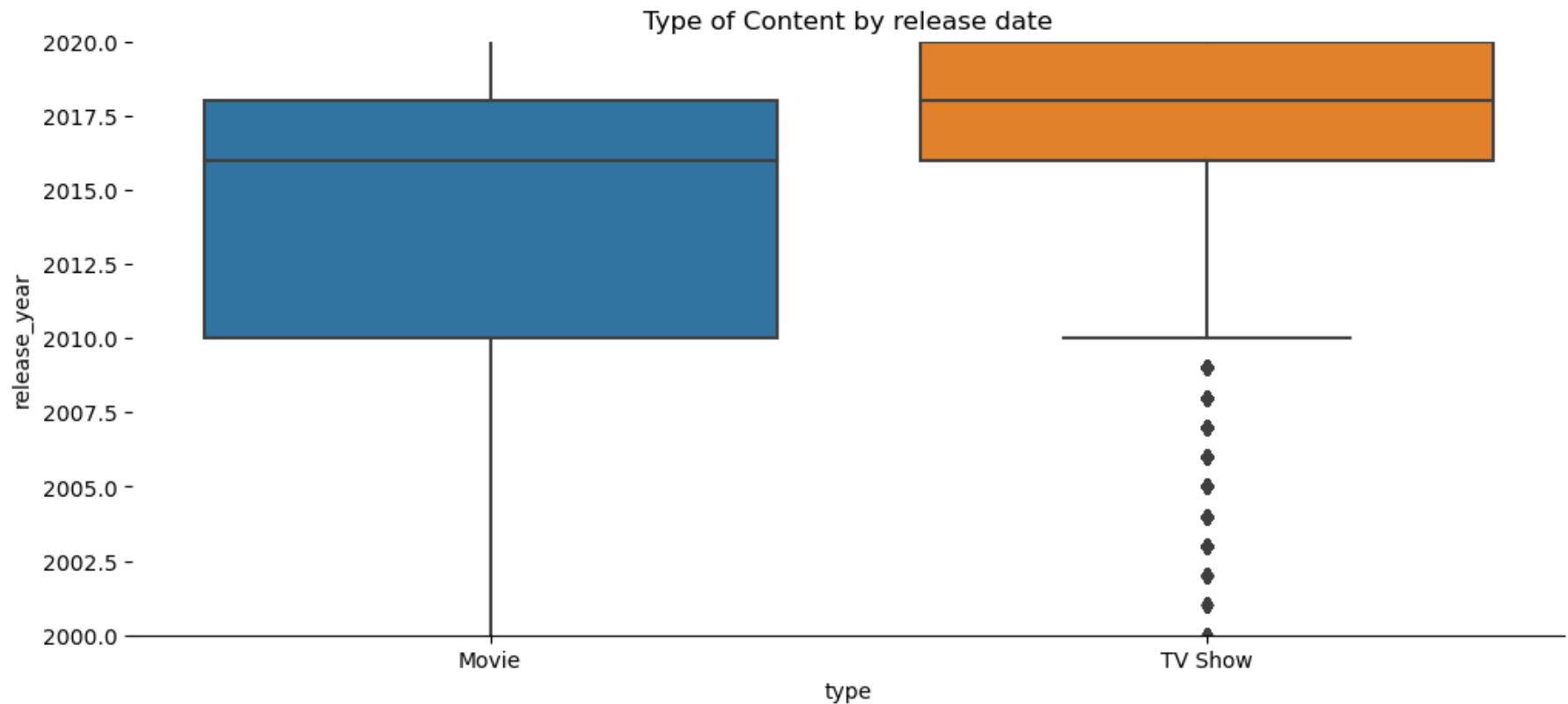


```
In [27]: # Checking for the amount of content added in the last ten years categorized by TV shows and Movies
# From this we can say that Netflix is focusing more on TV shows than movies from the past 5 years
plt.figure(figsize = (12,6))
sns.countplot(data = df, x = 'year', hue = 'type',palette=['#E50914',"#000000"])
plt.legend(loc = 'upper right',frameon = True, ncol = 2)
plt.show()
# from the past couple of years Netflix has been investing on TV shows as we can see some decent amount of incremen
# in tv shows being added into the netflix
```



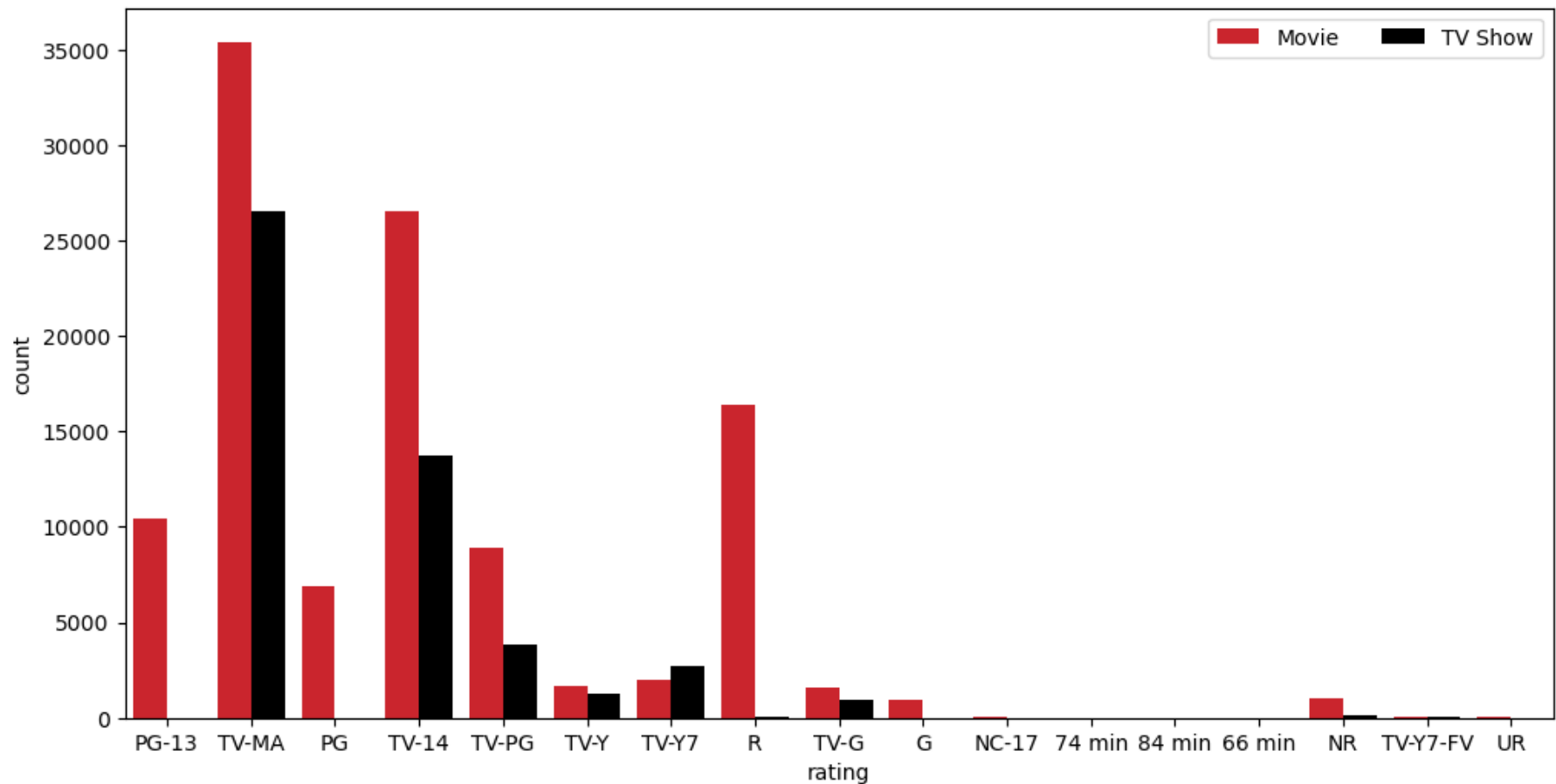
```
In [84]: # checking the content of recent releases ranging within a certain time period of time.  
# Analyzing the recent releases.
```

```
plt.figure(figsize=(12,5))  
sns.boxplot(x='type', y = 'release_year', data=df )  
sns.despine(left=True)  
plt.title('Type of Content by release date')  
plt.ylim(2000,2020)  
plt.show()
```



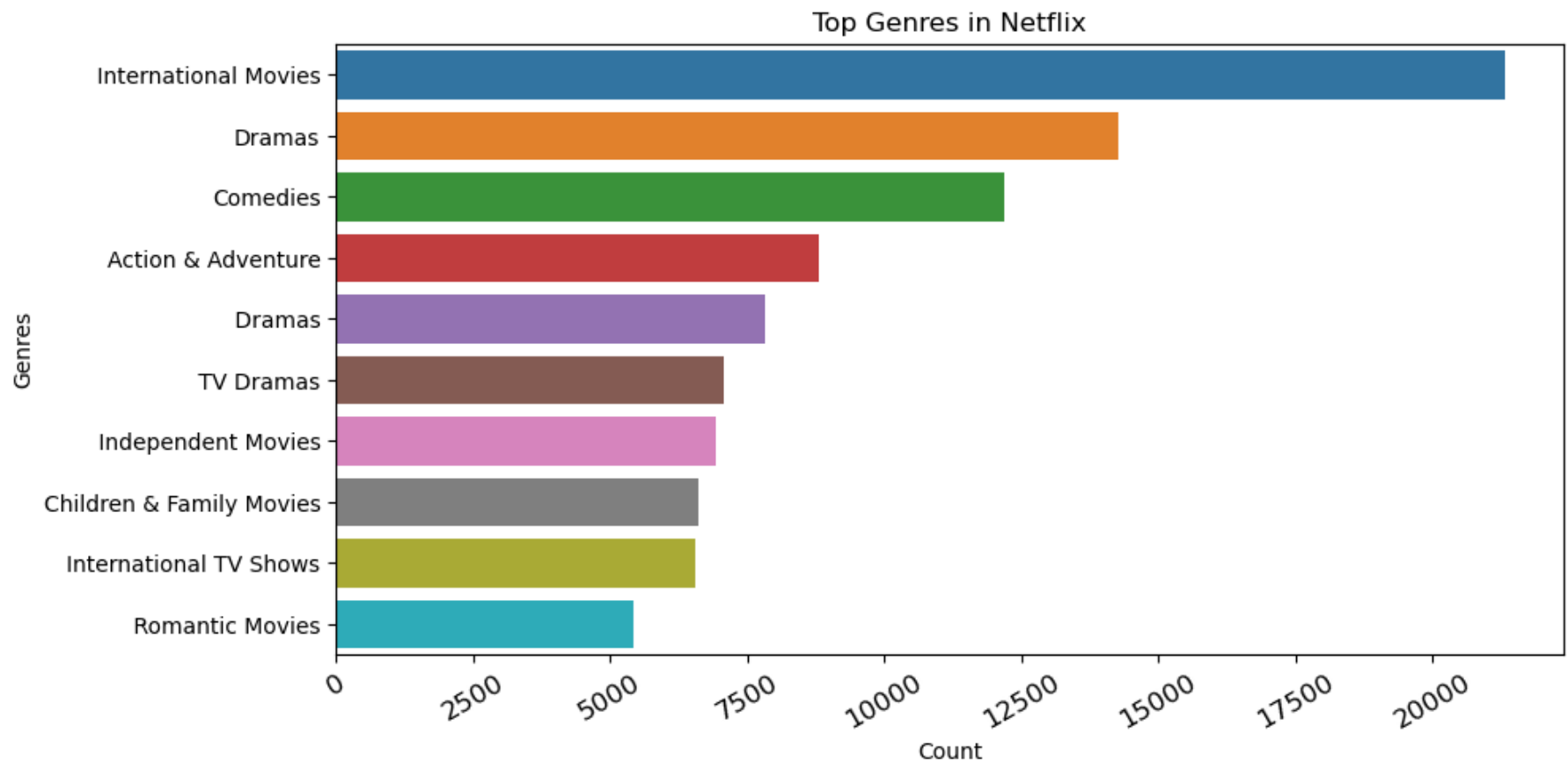
```
In [28]: # checking the ratings for content being streamed in Netflix for both Movies and TV shows  
# Visualization show cases the the ratings based upon the movies and tv shows in Netflix
```

```
plt.figure(figsize = (12,6))  
sns.countplot(data = df, x = 'rating', hue = 'type',palette=['#E50914',"#000000"])  
plt.legend(loc = 'upper right',frameon = True, ncol = 2)  
plt.show()
```

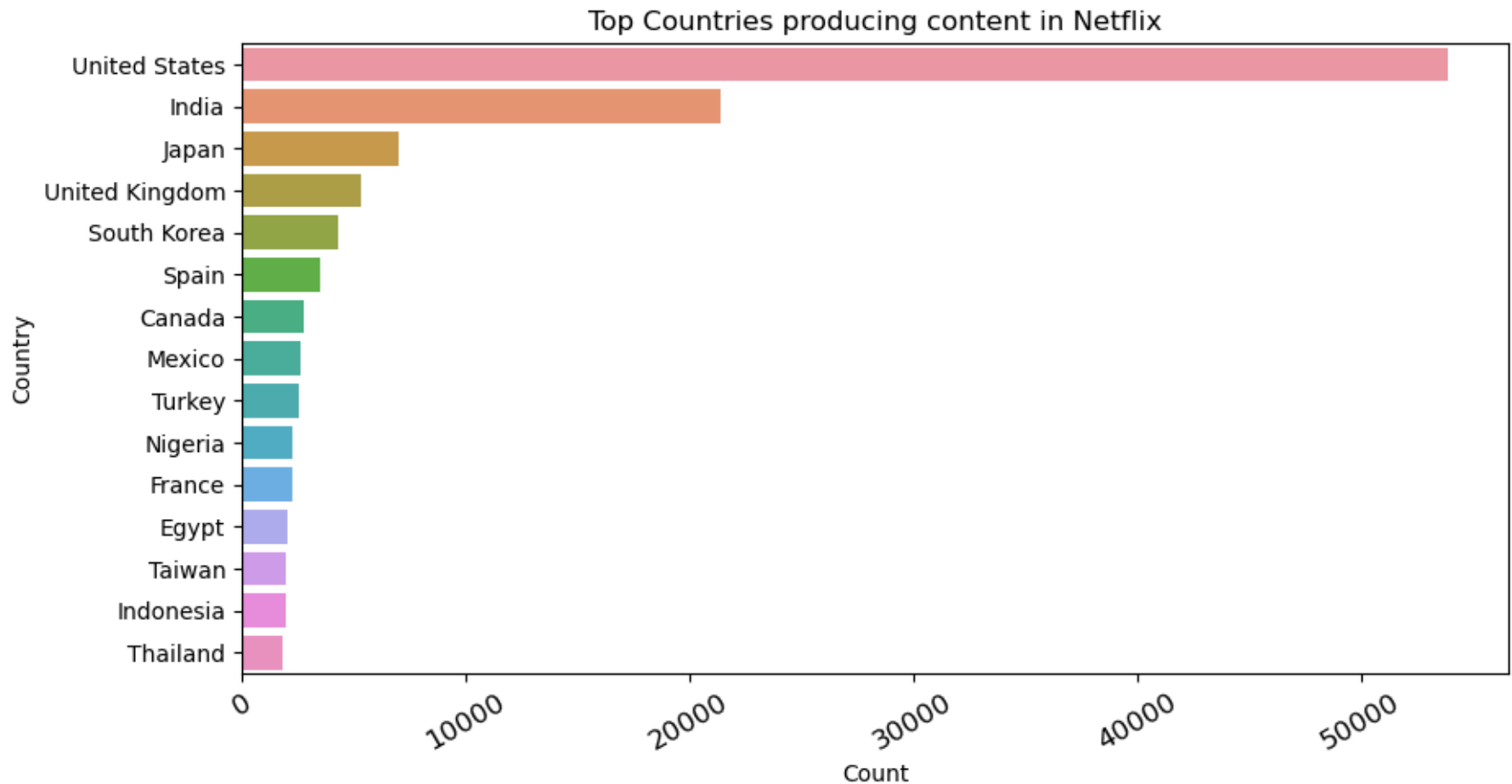




```
In [29]: # Top genres in Netflix based upon the amount of titles in the platform
plt.figure(figsize = (10,5))
sns.countplot(y = df['listed_in'],order = df['listed_in'].value_counts().index[:10])
plt.xticks(rotation = 30, fontsize = 12)
plt.title('Top Genres in Netflix')
plt.xlabel('Count')
plt.ylabel('Genres')
plt.show()
# Top genres in Netflix
```



```
In [30]: # content produced from different countries in the Netflix
plt.figure(figsize = (10,5))
sns.countplot(y = df['country'],order = df['country'].value_counts().index[:15])
plt.xticks(rotation = 30, fontsize = 12)
plt.title('Top Countries producing content in Netflix')
plt.xlabel('Count')
plt.ylabel('Country')
plt.show()
# Top content producers into Netflix
```



```
In [30]: # Filtering on the content of movies to understand the movies behaviour in Netflix  
movies_content = df.loc[df['type'] == 'Movie']
```

```
In [31]: movies_content
```

Out[31]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	year	mo
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Anupam Kher	United States	2021-09-25	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	2021	
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	United States	2021-09-24	2021	PG	91 min	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...	2021	
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	United States	2021-09-24	2021	PG	91 min	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...	2021	
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	James Marsden	United States	2021-09-24	2021	PG	91 min	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...	2021	
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Sofia Carson	United States	2021-09-24	2021	PG	91 min	Children & Family Movies	Equestria's divided. But a bright-eyed hero be...	2021	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2019-03-02	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...	2019	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2019-03-02	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...	2019	
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	111 min	Dramas	A scrappy but poor boy worms his way into a ty...	2019	

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	year	mo
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	111 min	International Movies	A scrappy but poor boy worms his way into a ty...	2019	
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	111 min	Music & Musicals	A scrappy but poor boy worms his way into a ty...	2019	

111989 rows × 14 columns

```
In [32]: # checking for the countries and amount of content they generate in terms of movies
movies_content['country'].value_counts()
```

```
Out[32]: United States      38259
India      20029
Japan      2486
United Kingdom      2165
Nigeria      2064
...
United States, Denmark      1
United States,      1
Netherlands, United States      1
Denmark, Sweden, Israel, United States      1
United States, Nicaragua      1
Name: country, Length: 651, dtype: int64
```

```
In [33]: # filtering out certain category of columns, based on the movies in Netflix
movie_content_in_countries = df.loc[df['type'] == 'Movie',['country','title','director','type','listed_in']]
```

In [34]: movie\_content\_in\_countries

Out[34]:

	country	title	director	type	listed_in
0	United States	Dick Johnson Is Dead	Kirsten Johnson	Movie	Documentaries
6	United States	My Little Pony: A New Generation	Robert Cullen	Movie	Children & Family Movies
6	United States	My Little Pony: A New Generation	Robert Cullen	Movie	Children & Family Movies
6	United States	My Little Pony: A New Generation	Robert Cullen	Movie	Children & Family Movies
6	United States	My Little Pony: A New Generation	Robert Cullen	Movie	Children & Family Movies
...	...	...	...	...	...
8806	India	Zubaan	Mozez Singh	Movie	International Movies
8806	India	Zubaan	Mozez Singh	Movie	Music & Musicals
8806	India	Zubaan	Mozez Singh	Movie	Dramas
8806	India	Zubaan	Mozez Singh	Movie	International Movies
8806	India	Zubaan	Mozez Singh	Movie	Music & Musicals

111989 rows × 5 columns

```
In [35]: # checking the genre count in movies  
# Among the movies content that has been generated in Netflix, checking for individual genre count.  
movie_content_in_countries['listed_in'].value_counts()
```

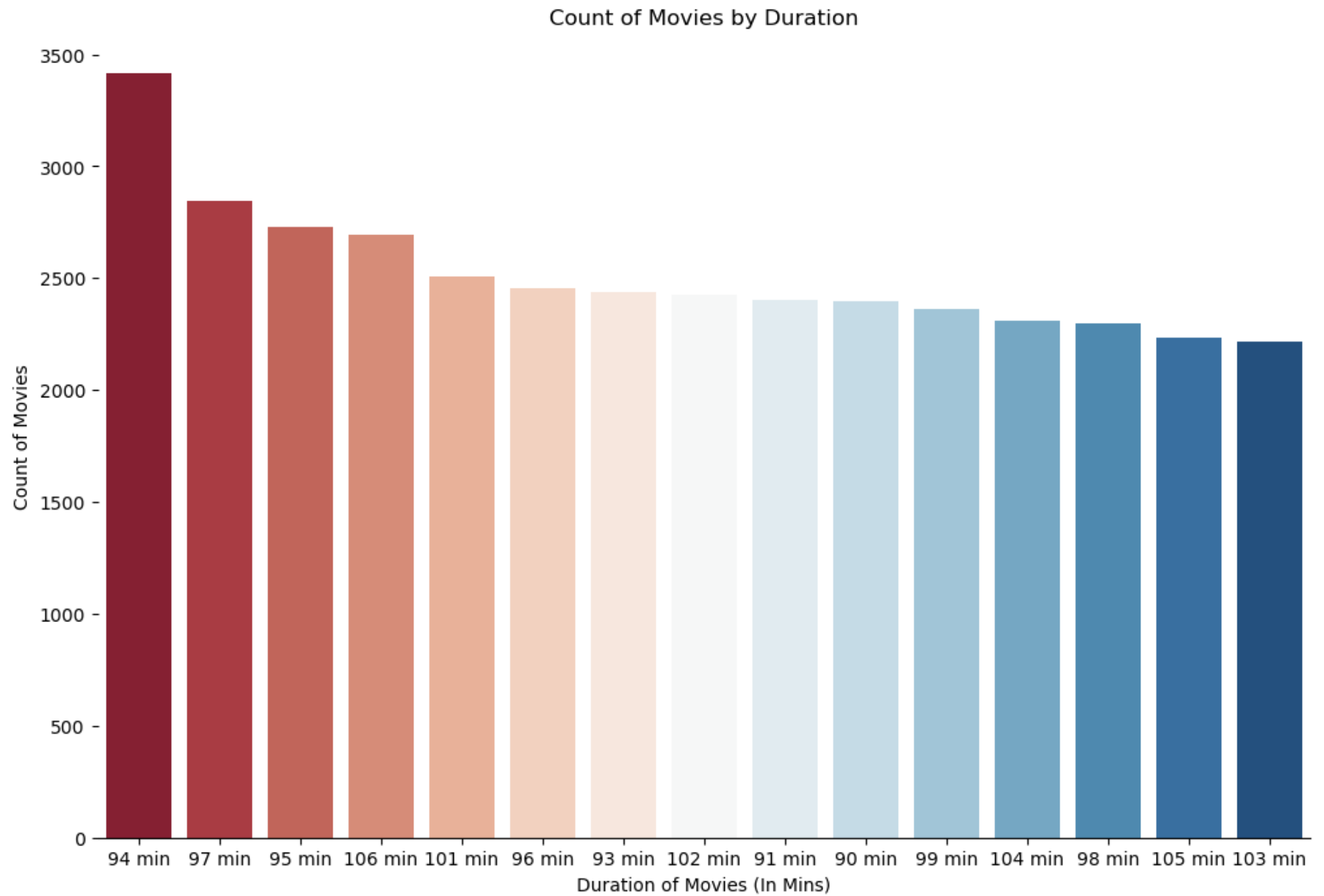


```

Out[35]: International Movies      21325
          Dramas                  14264
          Comedies                12184
          Action & Adventure      8805
          Dramas                  7833
          Independent Movies      6940
          Children & Family Movies 6607
          Romantic Movies        5434
          Comedies               5048
          Thrillers              4481
          Music & Musicals        2595
          Horror Movies          2557
          Sci-Fi & Fantasy        2496
          Documentaries          1819
          Sports Movies          1272
          International Movies    936
          Horror Movies          930
          LGBTQ Movies           722
          Classic Movies         697
          Anime Features         687
          Cult Movies            656
          Thrillers             597
          Faith & Spirituality    531
          Stand-Up Comedy        515
          Children & Family Movies 411
          Classic Movies         388
          Movies                 371
          Anime Features         241
          Independent Movies     186
          Cult Movies            125
          Documentaries          118
          Sci-Fi & Fantasy        107
          Music & Musicals        59
          Stand-Up Comedy        24
          Romantic Movies        20
          LGBTQ Movies           5
          Sports Movies          3
          Name: listed_in, dtype: int64

```

```
In [157]: # checking for the duration values that has been associated with movies in Netflix
df_md = movies_content.groupby('duration').count().reset_index().sort_values(ascending=False,by='type')[0:15]
plt.figure(figsize=(12,8))
sns.barplot(data = df_md, x='duration', y='show_id', order=df_md['duration'], palette='RdBu')
sns.despine(left=True)
plt.ylabel('Count of Movies')
plt.xlabel('Duration of Movies (In Mins)')
plt.title('Count of Movies by Duration')
plt.show()
```



```
In [47]: # filtering the data based upon tv shows in netflix to understand their behaviour  
tv_show_content_in_countries = df.loc[df['type'] == 'TV Show']
```

```
In [48]: tv_show_content_in_countries
```

Out[48]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	year	month
1	s2	TV Show	Blood & Water	Cathy Garcia-Molina	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...	2021	09
1	s2	TV Show	Blood & Water	Cathy Garcia-Molina	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...	2021	09
1	s2	TV Show	Blood & Water	Cathy Garcia-Molina	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Mysteries	After crossing paths at a party, a Cape Town t...	2021	09
1	s2	TV Show	Blood & Water	Cathy Garcia-Molina	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...	2021	09
1	s2	TV Show	Blood & Water	Cathy Garcia-Molina	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...	2021	09
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8800	s8801	TV Show	Zindagi Gulzar Hai	Cathy Garcia-Molina	Hina Khawaja Bayat	Pakistan	2016-12-15	2012	TV-PG	1 Season	Romantic TV Shows	Strong-willed, middle-class Kashaf and carefre...	2016	12
8800	s8801	TV Show	Zindagi Gulzar Hai	Cathy Garcia-Molina	Hina Khawaja Bayat	Pakistan	2016-12-15	2012	TV-PG	1 Season	TV Dramas	Strong-willed, middle-class Kashaf and carefre...	2016	12
8803	s8804	TV Show	Zombie Dumb	Cathy Garcia-Molina	Anupam Kher	United States	2019-07-01	2018	TV-Y7	2 Seasons	Kids' TV	While living alone in a spooky town, a young g...	2019	07
8803	s8804	TV Show	Zombie Dumb	Cathy Garcia-Molina	Anupam Kher	United States	2019-07-01	2018	TV-Y7	2 Seasons	Korean TV Shows	While living alone in a spooky town, a young g...	2019	07

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	year	month
8803	s8804	TV Show	Zombie Dumb	Cathy Garcia-Molina	Anupam Kher	United States	2019-07-01	2018	TV-Y7	2 Seasons	TV Comedies	While living alone in a spooky town, a young g...	2019	07

49227 rows × 14 columns

```
In [49]: # checking for genre count in tv shows
tv_show_content_in_countries['listed_in'].value_counts()
```

```
Out[49]: TV Dramas          7070
International TV Shows    6564
International TV Shows    5154
Crime TV Shows           3449
TV Comedies              3422
Kids' TV                 2763
Romantic TV Shows        2650
Anime Series             2126
Spanish-Language TV Shows 1825
British TV Shows         1434
TV Action & Adventure     1331
TV Comedies              1167
TV Mysteries             1138
Korean TV Shows          1086
TV Sci-Fi & Fantasy        859
TV Dramas                748
Teen TV Shows            742
TV Horror                 661
TV Thrillers             650
Crime TV Shows           607
Kids' TV                 505
Docuseries               479
TV Action & Adventure     465
TV Shows                 337
Reality TV               293
Romantic TV Shows        290
Reality TV               285
Docuseries               274
Classic & Cult TV         204
Stand-Up Comedy & Talk Shows 188
TV Horror                180
Science & Nature TV       130
Stand-Up Comedy & Talk Shows 80
Classic & Cult TV         36
Spanish-Language TV Shows 28
TV Sci-Fi & Fantasy        7
Name: listed_in, dtype: int64
```



```
In [50]: # checking content of tv shows and the countries producing them
tv_show_content_in_countries['country'].value_counts()
```

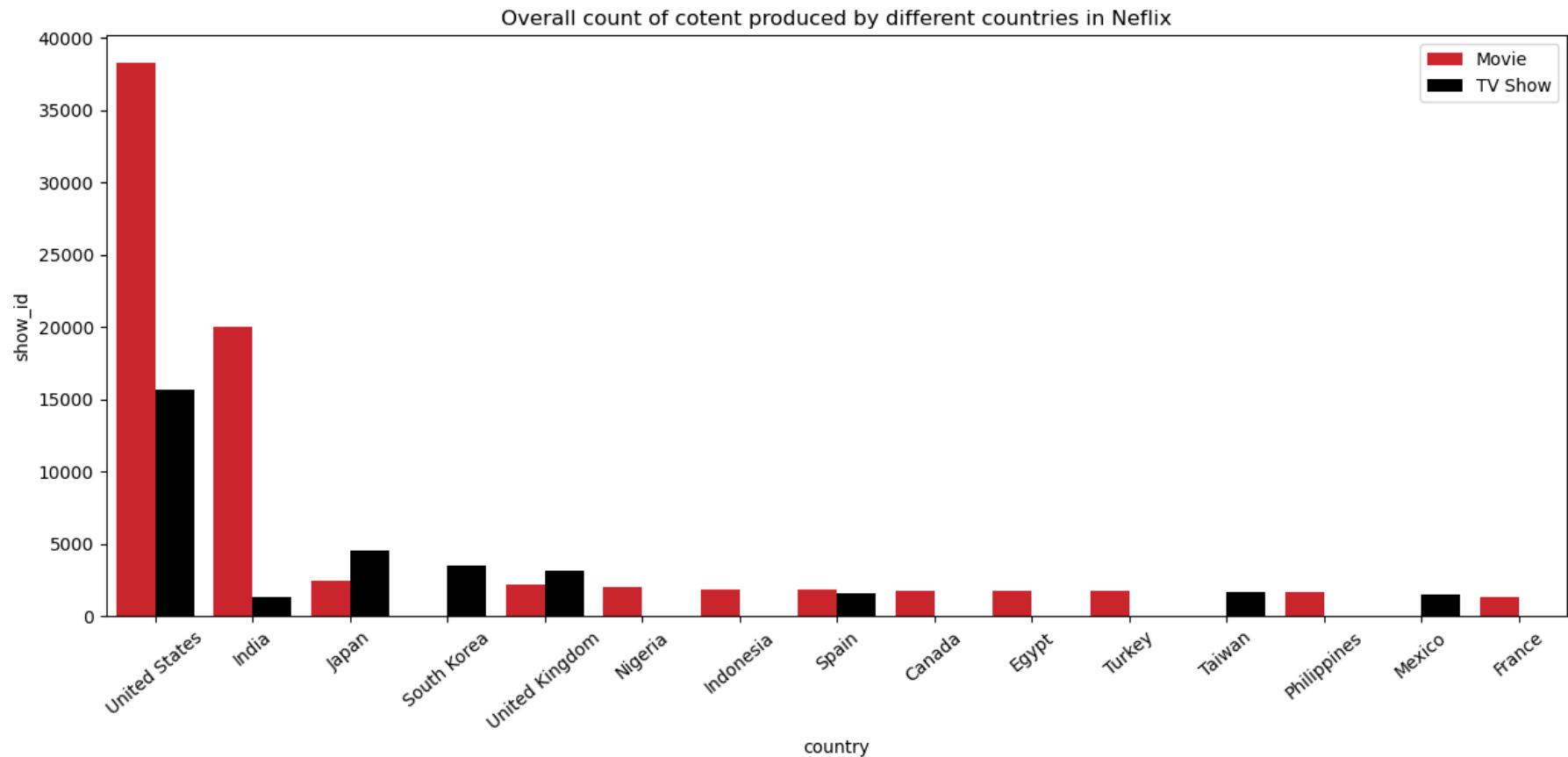
```
Out[50]: United States          15640
         Japan                  4540
         South Korea            3531
         United Kingdom         3188
         Taiwan                 1675
         ...
         Brazil, United Kingdom    2
         Argentina, United States, Mexico    2
         France, United Kingdom, United States    1
         United States, Russia    1
         Czech Republic, United States    1
         Name: country, Length: 196, dtype: int64
```

```
In [51]: # checking for director data, that are part of creating tv shows
tv_show_content_in_countries['director'].value_counts()
```

```
Out[51]: Cathy Garcia-Molina    43417
         Rob Seidenglanz        103
         Kongkiat Komesiri      75
         Danny Cannon           75
         Pass Patthanakumjon     66
         ...
         Julia Reichert          1
         Steven Bognar           1
         Sharon Grimberg         1
         Hikaru Toda             1
         Ronna Gradus            1
         Name: director, Length: 301, dtype: int64
```

In [52]: *# visualizing the overall content of movies and tv shows produced by countries in the netflix till data.*

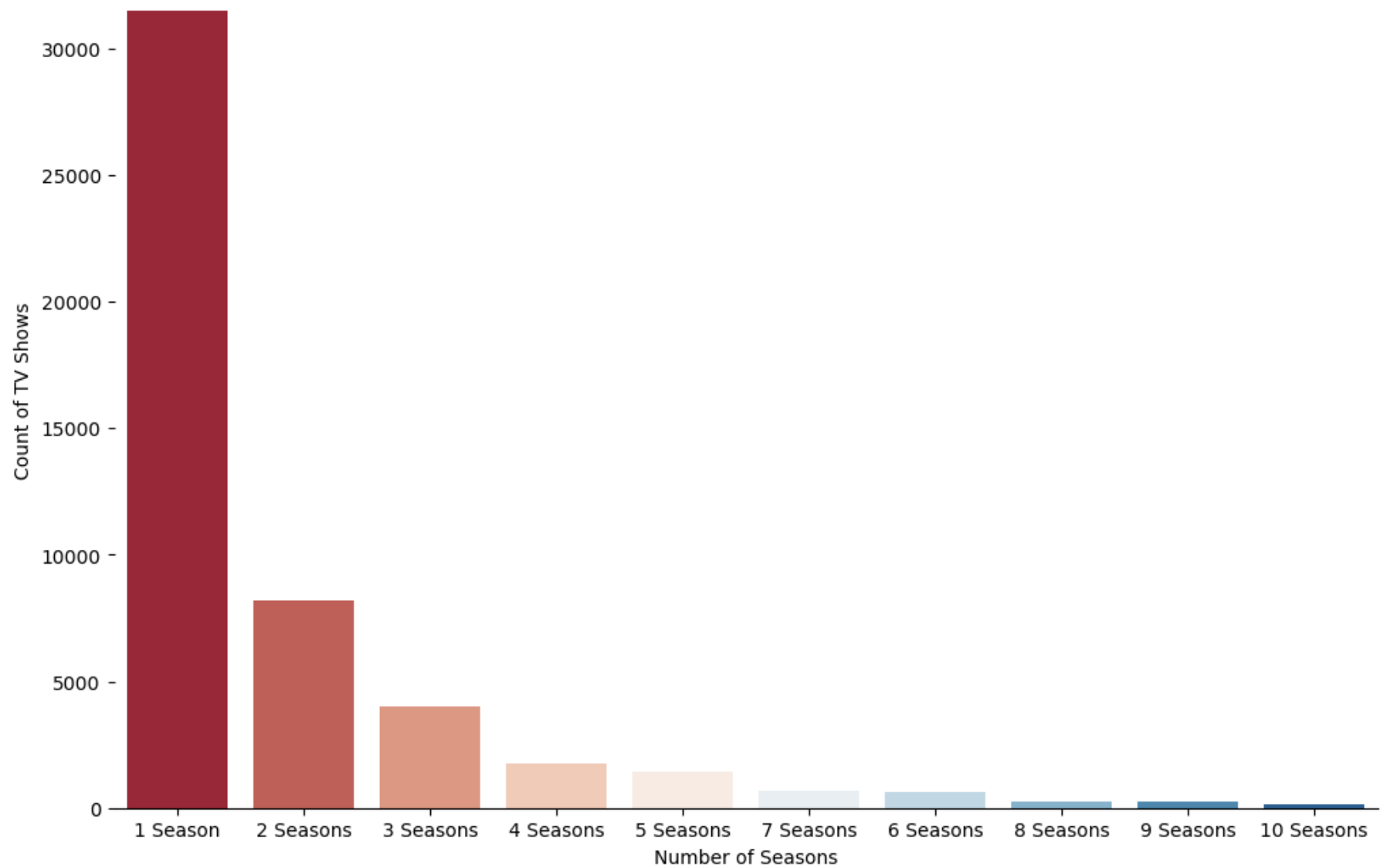
```
df_country = df.groupby(['country', 'type'])['show_id'].count().sort_values(ascending = False).reset_index()
plt.figure(figsize = (15, 6))
sns.barplot(data = df_country, x = df_country['country'][:20], y = df_country['show_id'], hue = 'type',
            palette=['#E50914', "#000000"])
plt.xticks(rotation = 40)
plt.legend(loc = 'upper right')
plt.title('Overall count of cotent produced by different countries in Neflix')
plt.show()
```



In [55]: *# count of tv shows with duration in seasons*

```
df_td = tv_show_content_in_countries.groupby('duration').count().reset_index().sort_values(ascending=False,by='type')
plt.figure(figsize=(12,8))
sns.barplot(data = df_td, x='duration', y='show_id', order=df_td['duration'], palette='RdBu')
sns.despine(left=True)
plt.ylabel('Count of TV Shows')
plt.xlabel('Number of Seasons')
plt.title('Count of Tv Shows by Number of Seasons')
plt.show()
```

Count of Tv Shows by Number of Seasons



In [56]: *# checking for the trends, trying to find a correlation between duration and content over the years*

```
df_mr = df[df['type'] == 'Movie'].groupby('release_year').count().reset_index().sort_values(ascending=False, by='type')
df_tr = df[df['type'] == 'TV Show'].groupby('release_year').count().reset_index().sort_values(ascending=False, by='type')
```

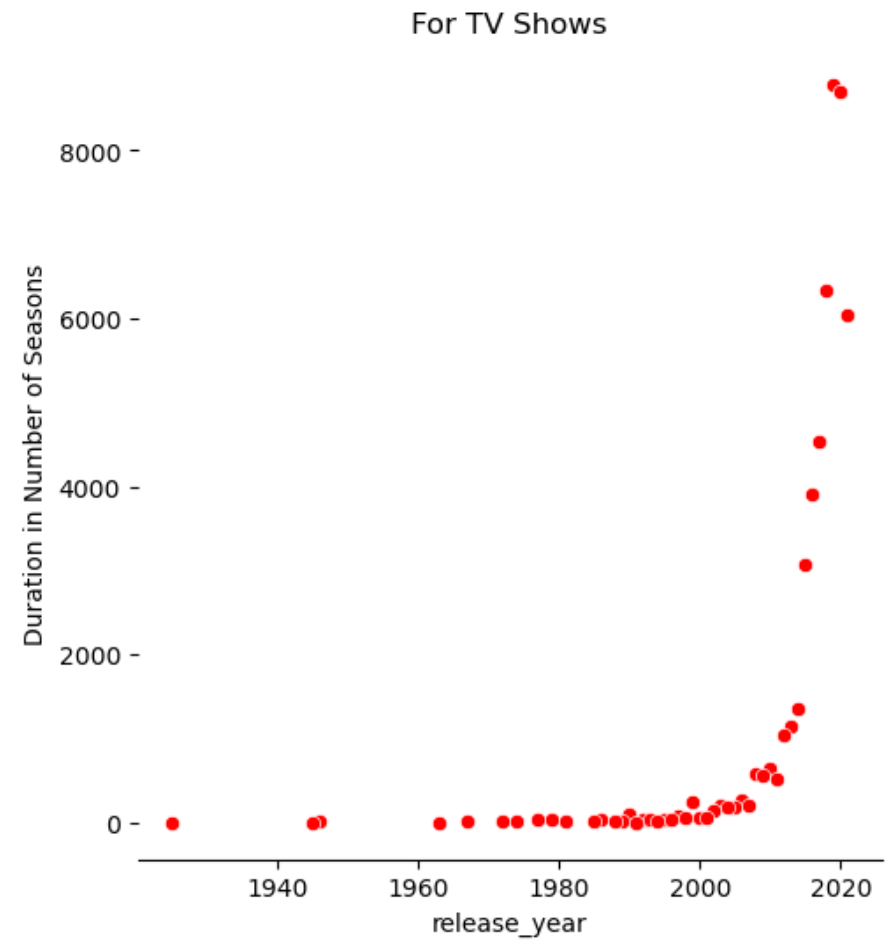
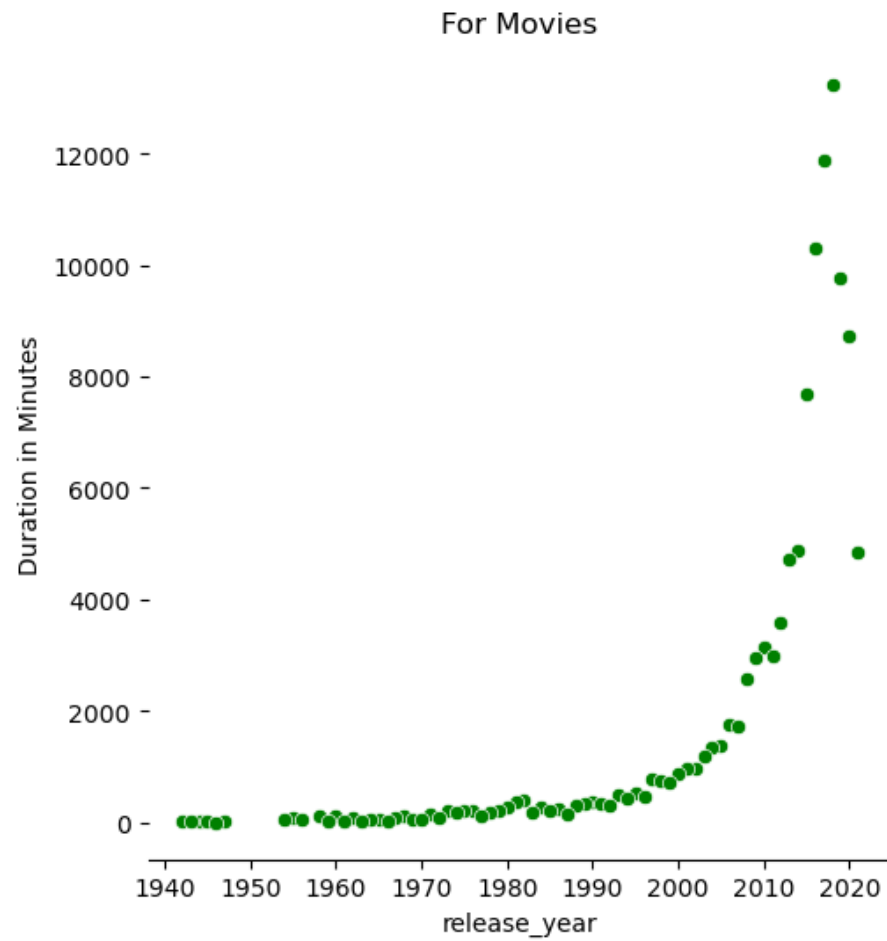
In [58]: *# visual representation of correlation between content and duration.*

```
fig, ax = plt.subplots(nrows= 1, ncols = 2, figsize=(12,6))
g = sns.scatterplot(x='release_year', y='duration', data=df_mr, ax=ax[0], color= 'green')
s = sns.scatterplot(x='release_year', y='duration', data=df_tr, ax=ax[1], color='red')
sns.despine(left=True)

g.set_ylabel('Duration in Minutes')
s.set_ylabel('Duration in Number of Seasons')

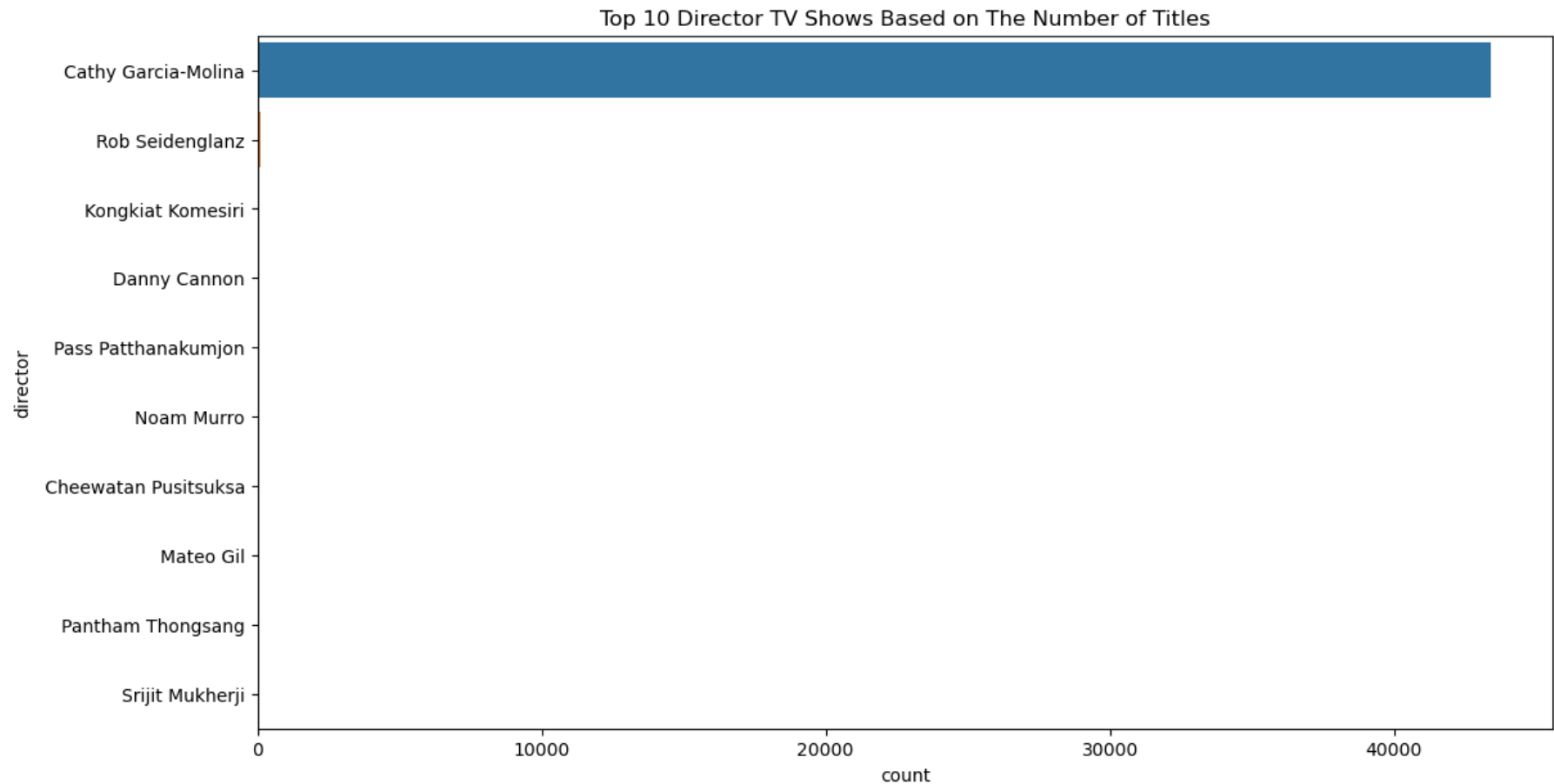
g.set_title('For Movies')
s.set_title('For TV Shows')
plt.show()

# considerably the duration in content is getting increased as the years are progressing
```



In [123]: *# checking for Top 10 Director tv shows based on number of titles in Netflix.*

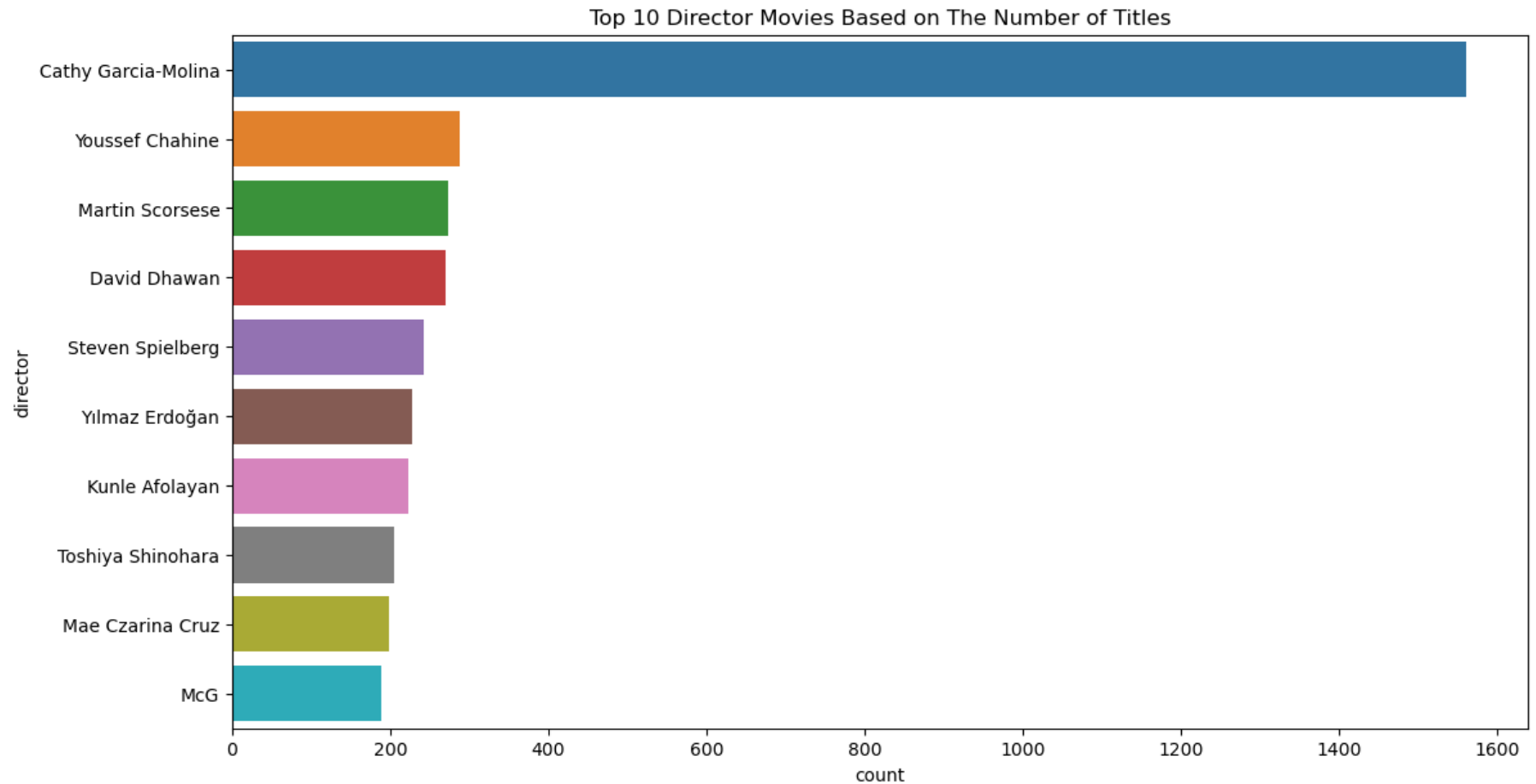
```
plt.figure(figsize=(13,7))  
plt.title('Top 10 Director TV Shows Based on The Number of Titles')  
sns.countplot(y = tv_show_content_in_countries['director'], order=tv_show_content_in_countries['director'].value_count  
plt.show()
```





In [124]: *# checking for Top 10 Director movies based on number of titles in Netflix*

```
plt.figure(figsize=(13,7))  
plt.title('Top 10 Director Movies Based on The Number of Titles')  
sns.countplot(y = movie_content_in_countries['director'], order=movie_content_in_countries['director'].value_counts()).  
plt.show()
```



```
In [87]: # Analyzing the data based on category and highly rated by audience  
# Analysis is being done on both actor and director in the further steps.  
  
tv_show_data = df.loc[(df['type'] == 'TV Show') & (df['rating'] == 'TV-MA')]  
movies_data = df.loc[(df['type'] == 'Movie') & (df['rating'] == 'TV-MA')]
```

In [88]: tv\_show\_data

Out[88]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	year	month
1	s2	TV Show	Blood & Water	Cathy Garcia-Molina	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...	2021	9
1	s2	TV Show	Blood & Water	Cathy Garcia-Molina	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...	2021	9
1	s2	TV Show	Blood & Water	Cathy Garcia-Molina	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Mysteries	After crossing paths at a party, a Cape Town t...	2021	9
1	s2	TV Show	Blood & Water	Cathy Garcia-Molina	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t...	2021	9
1	s2	TV Show	Blood & Water	Cathy Garcia-Molina	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t...	2021	9
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8749	s8750	TV Show	Winter Sun	Cathy Garcia-Molina	Gamze Süner Atay	Turkey	2017-08-01	2014	TV-MA	1 Season	TV Dramas	Years after ruthless businessmen kill his fath...	2017	8
8749	s8750	TV Show	Winter Sun	Cathy Garcia-Molina	Gamze Süner Atay	Turkey	2017-08-01	2014	TV-MA	1 Season	TV Thrillers	Years after ruthless businessmen kill his fath...	2017	8
8749	s8750	TV Show	Winter Sun	Cathy Garcia-Molina	Mehmet Esen	Turkey	2017-08-01	2014	TV-MA	1 Season	International TV Shows	Years after ruthless businessmen kill his fath...	2017	8
8749	s8750	TV Show	Winter Sun	Cathy Garcia-Molina	Mehmet Esen	Turkey	2017-08-01	2014	TV-MA	1 Season	TV Dramas	Years after ruthless businessmen kill his fath...	2017	8

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	year	month
8749	s8750	TV Show	Winter Sun	Cathy Garcia-Molina	Mehmet Esen	Turkey	2017-08-01	2014	TV-MA	1 Season	TV Thrillers	Years after ruthless businessmen kill his fath...	2017	8

26546 rows × 14 columns

```
In [158]: # most liked content in tv_shows
# upcoming few steps will be more of descriptive anaylsis based on tv show content
most_liked_content_in_tv_shows = tv_show_data.loc[tv_show_data['rating'] == 'TV-MA', ['listed_in', 'cast', 'director', 'y
```

```
In [159]: most_liked_content_in_tv_shows
```

```
Out[159]:
```

	listed_in	cast	director	year
1	International TV Shows	Ama Qamata	Cathy Garcia-Molina	2021
1	TV Dramas	Ama Qamata	Cathy Garcia-Molina	2021
1	TV Mysteries	Ama Qamata	Cathy Garcia-Molina	2021
1	International TV Shows	Khosi Ngema	Cathy Garcia-Molina	2021
1	TV Dramas	Khosi Ngema	Cathy Garcia-Molina	2021
...	...	...	...	...
8749	TV Dramas	Gamze Süner Atay	Cathy Garcia-Molina	2017
8749	TV Thrillers	Gamze Süner Atay	Cathy Garcia-Molina	2017
8749	International TV Shows	Mehmet Esen	Cathy Garcia-Molina	2017
8749	TV Dramas	Mehmet Esen	Cathy Garcia-Molina	2017
8749	TV Thrillers	Mehmet Esen	Cathy Garcia-Molina	2017

26546 rows × 4 columns

```
In [160]: # gathering actor analysis who acted in tv shows that has been highly rated
most_liked_content_in_tv_shows['cast'].value_counts()
```

```
Out[160]: Anupam Kher          357
          Hsia Teng-hung      32
          Takahiro Sakurai    30
          River Huang         27
          Janel Tsai          24
          ...
          Tony Hale           1
          Will Arnett          1
          Portia de Rossi      1
          Aramis Knight        1
          Boulet Brothers      1
          Name: cast, Length: 8332, dtype: int64
```

```
In [161]: # gathering director analysis who directed tv shows that has been highly rated
most_liked_content_in_tv_shows['director'].value_counts()
```

```
Out[161]: Cathy Garcia-Molina  22555
          Kongkiat Komesiri     75
          Danny Cannon          75
          Pass Patthanakumjon   66
          Mateo Gil              60
          ...
          Eric Abrams            1
          Rashida Jones          1
          Ken Burns              1
          Lynn Novick            1
          Jill Bauer             1
          Name: director, Length: 182, dtype: int64
```

```
In [162]: # gathering the genre and its count of movies in netflix which has been highly rated by audience.  
most_liked_content_in_tv_shows['listed_in'].value_counts()
```

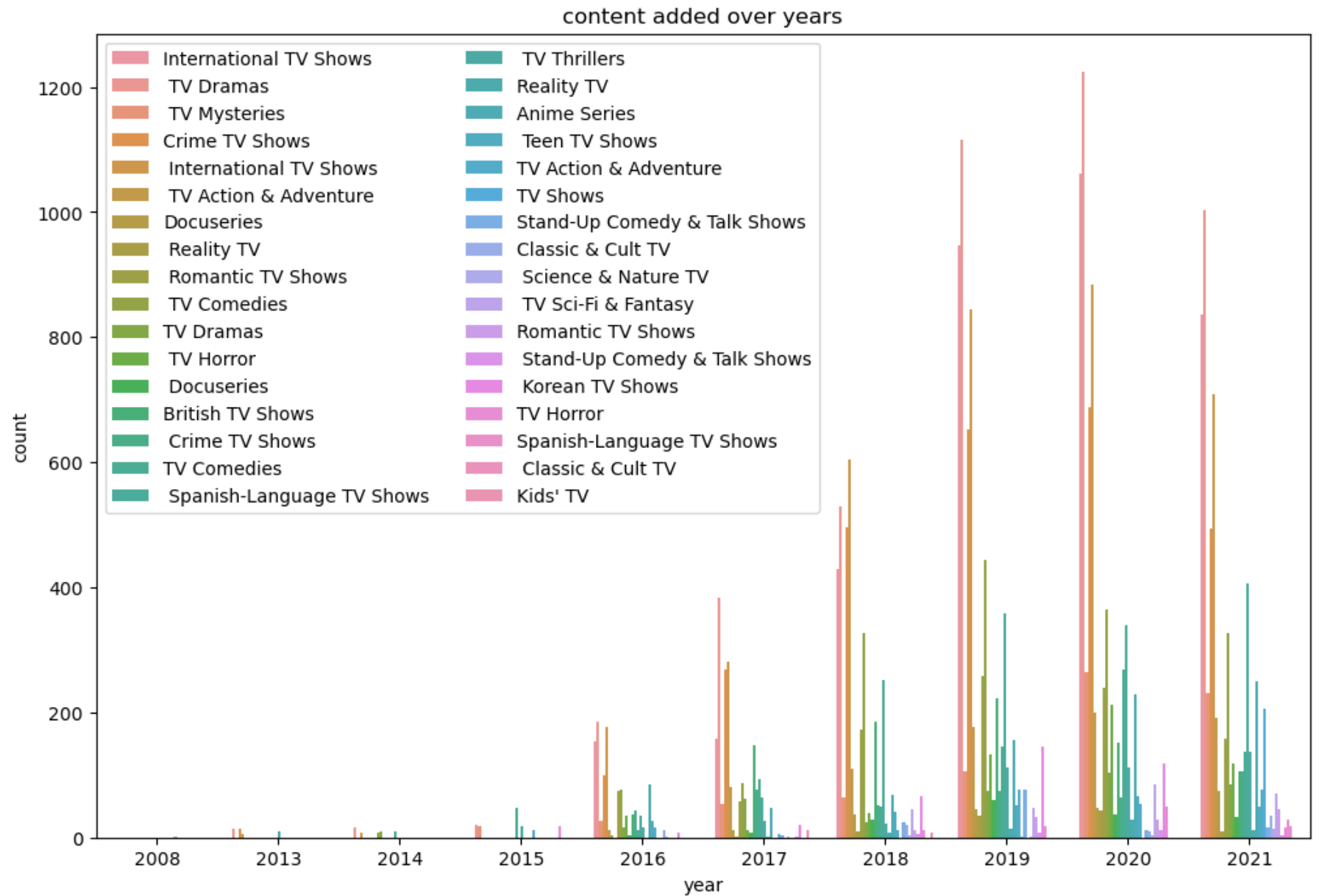
```
Out[162]: TV Dramas 4494  
International TV Shows 3585  
International TV Shows 3503  
Crime TV Shows 2719  
TV Comedies 1632  
Spanish-Language TV Shows 1455  
Romantic TV Shows 959  
British TV Shows 849  
Anime Series 834  
TV Action & Adventure 768  
TV Comedies 765  
TV Mysteries 764  
TV Horror 550  
TV Thrillers 458  
Crime TV Shows 418  
Korean TV Shows 375  
TV Dramas 374  
TV Sci-Fi & Fantasy 250  
TV Action & Adventure 246  
Teen TV Shows 234  
Docuseries 220  
TV Shows 214  
Docuseries 170  
Stand-Up Comedy & Talk Shows 136  
TV Horror 126  
Romantic TV Shows 119  
Reality TV 100  
Classic & Cult TV 76  
Reality TV 61  
Stand-Up Comedy & Talk Shows 30  
Science & Nature TV 26  
Spanish-Language TV Shows 18  
Classic & Cult TV 11  
Kids' TV 7  
Name: listed_in, dtype: int64
```

In [163]:

```
# highest rated content across tv shows over the course of years.
```

```
plt.figure(figsize = (12,8))  
sns.countplot(data = most_liked_content_in_tv_shows, x = 'year', hue = 'listed_in')  
plt.legend(loc = 'upper left', frameon = True, ncol = 2)  
plt.title('content added over years')  
plt.show()
```





```
In [97]: # most liked content in movies
# upcoming few steps will be more of descriptive anaylsis based on movie content
most_liked_content_in_movies = movies_data.loc[movies_data['rating'] == 'TV-MA', ['listed_in', 'cast', 'director', 'year']
```

```
In [294]: most_liked_content_in_movies
```

```
Out[294]:
```

	listed_in	cast	director	year
7	Dramas	Kofi Ghanaba	Haile Gerima	2021
7	Independent Movies	Kofi Ghanaba	Haile Gerima	2021
7	International Movies	Kofi Ghanaba	Haile Gerima	2021
7	Dramas	Oyafunmike Ogunlano	Haile Gerima	2021
7	Independent Movies	Oyafunmike Ogunlano	Haile Gerima	2021
...	...	...	...	...
8801	International Movies	Mansoor Alfeeli	Majid Al Ansari	2016
8801	Thrillers	Mansoor Alfeeli	Majid Al Ansari	2016
8801	Dramas	Ahd	Majid Al Ansari	2016
8801	International Movies	Ahd	Majid Al Ansari	2016
8801	Thrillers	Ahd	Majid Al Ansari	2016

35410 rows × 4 columns

```
In [98]: # gathering the genre and its count of movies in netflix which has been highly rated by audience.
most_liked_content_in_movies['listed_in'].value_counts()
```

```
Out[98]: International Movies      9216
Dramas                          5553
Comedies                        4068
Independent Movies              3159
Dramas                         2092
Thrillers                      1987
Action & Adventure             1926
Romantic Movies                1285
Horror Movies                  1273
Documentaries                  760
Music & Musicals                640
Stand-Up Comedy                445
Sci-Fi & Fantasy                436
LGBTQ Movies                   415
International Movies           403
Sports Movies                  356
Horror Movies                  328
Comedies                       292
Anime Features                 213
Movies                         93
Independent Movies             92
Classic Movies                 85
Thrillers                      59
Cult Movies                    46
Faith & Spirituality           39
Documentaries                  33
Music & Musicals               23
Stand-Up Comedy               21
Children & Family Movies       21
Sci-Fi & Fantasy               21
Classic Movies                 10
Anime Features                 10
Romantic Movies                10
Name: listed_in, dtype: int64
```

```
In [295]: # gathering actor analysis who acted in movies that has been highly rated
most_liked_content_in_movies['cast'].value_counts()
```

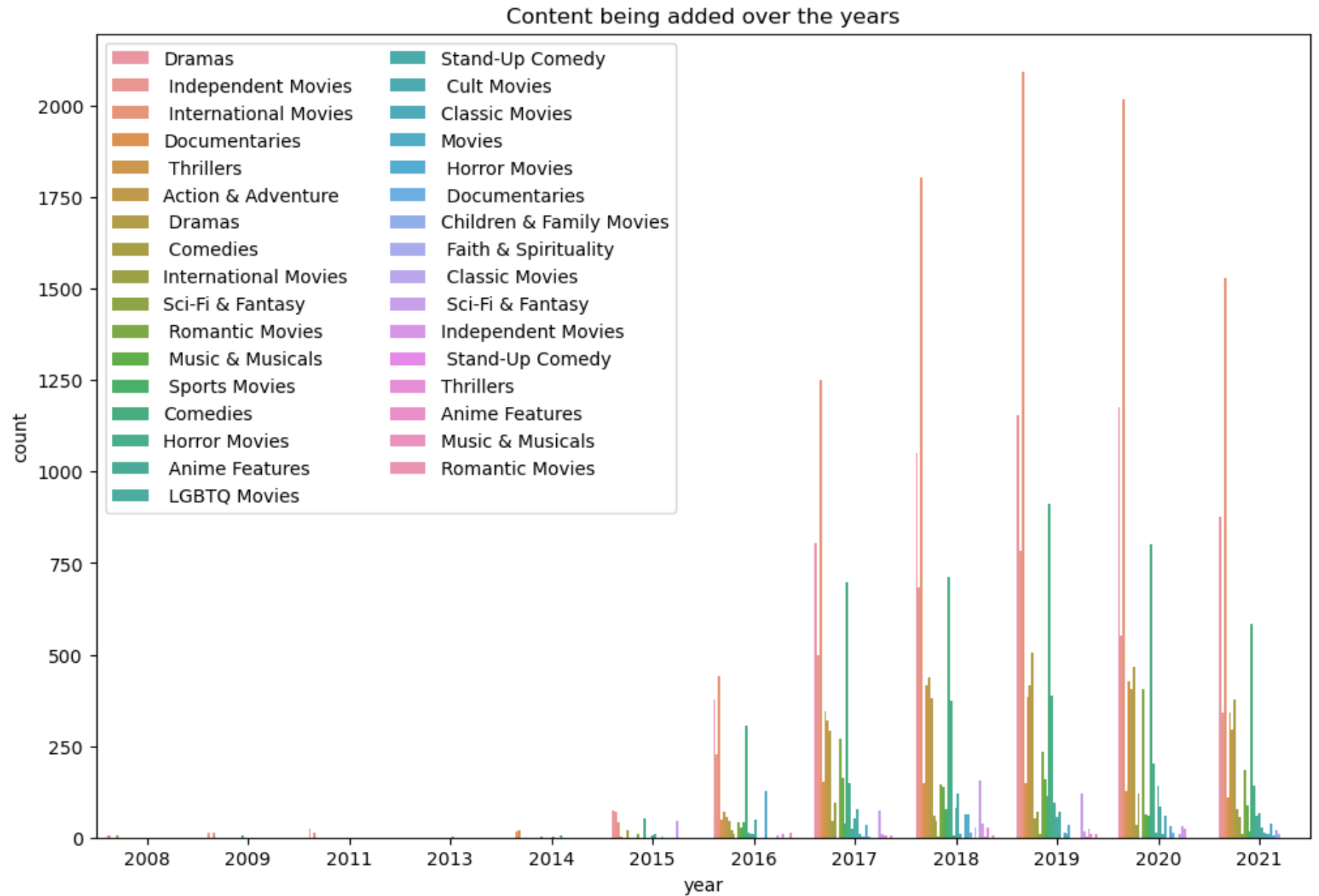
```
Out[295]: Anupam Kher          399
          Radhika Apte        59
          Piaa Bajpai         36
          Usha Uthup          36
          Rajat Kapoor        35
          ...
          Omar Wasow          1
          Jessica Taylor-Bearman 1
          Fortune Feimster     1
          Jackson Nicoll       1
          Iain Glen            1
          Name: cast, Length: 10996, dtype: int64
```

```
In [296]: # gathering director analysis who directed movies that has been highly rated
most_liked_content_in_movies['director'].value_counts()
```

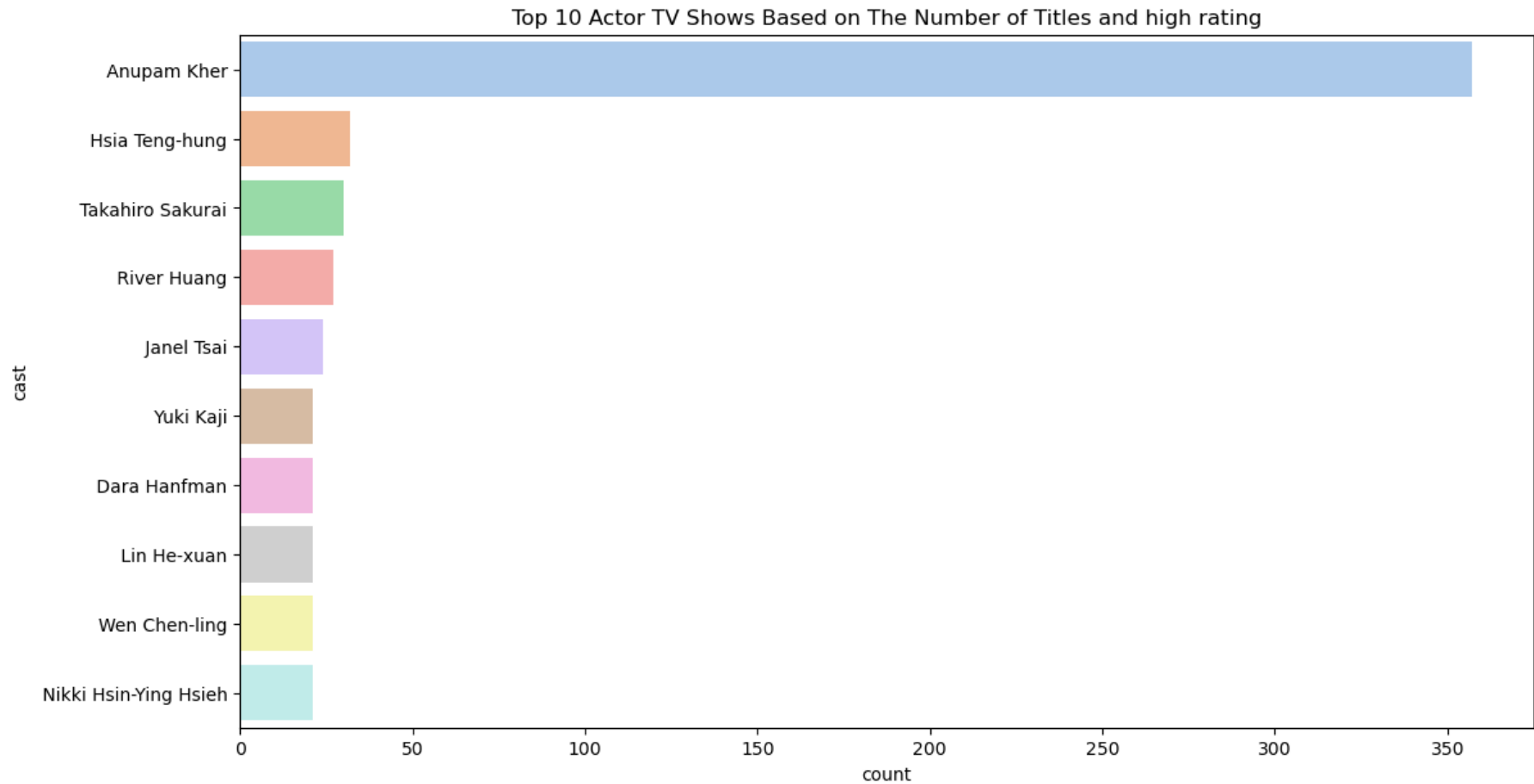
```
Out[296]: Cathy Garcia-Molina  455
          Youssef Chahine      179
          Shigeaki Kubo        160
          Akiva Schaffer       159
          Scott Aukerman       128
          ...
          Chris Rock           1
          Dave Higby           1
          Jonathan Demme       1
          Kevin MacDonald      1
          August Jakobsson     1
          Name: director, Length: 1910, dtype: int64
```

In [165]: *# Highly rated movies content over course of years*

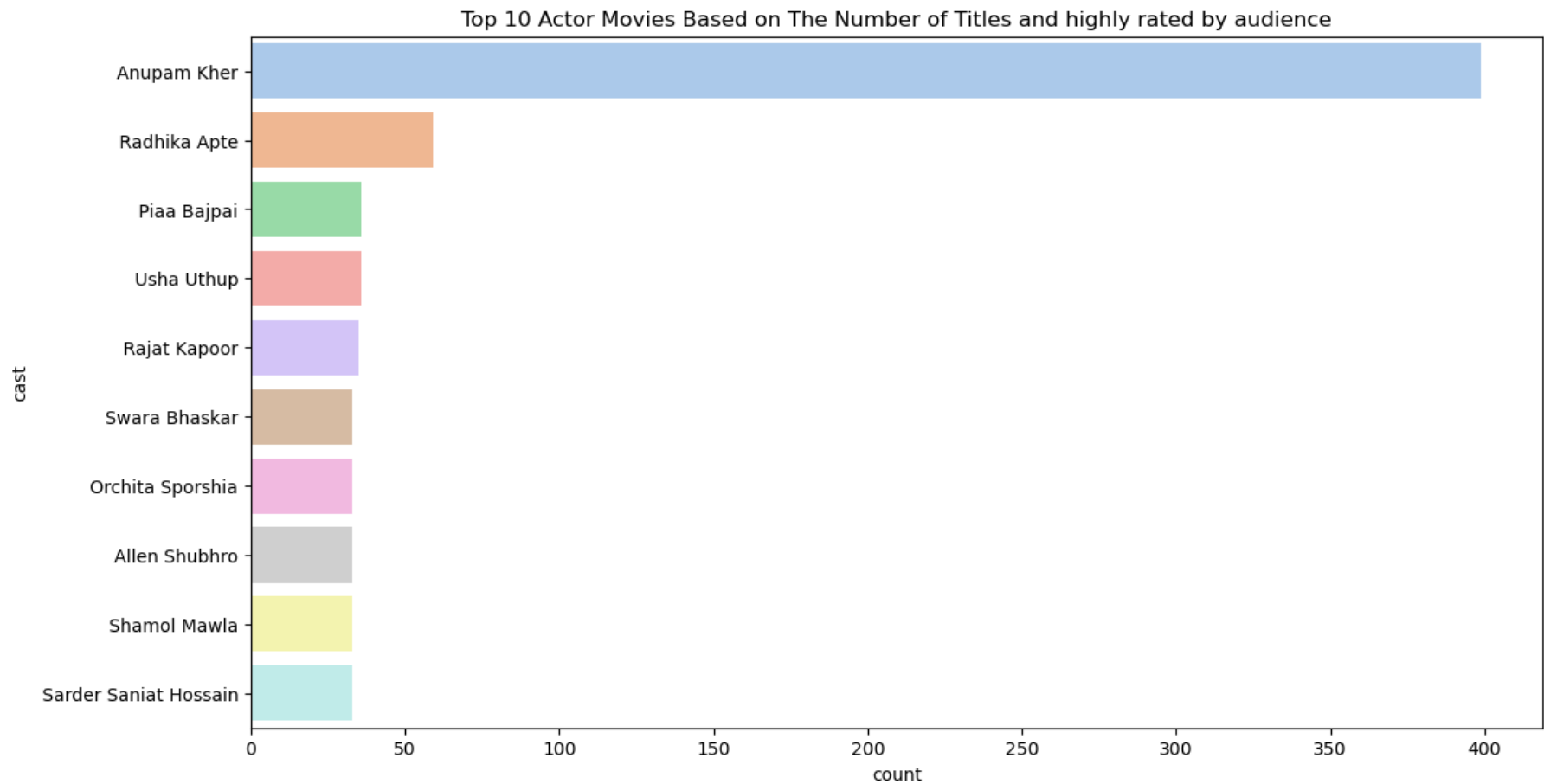
```
plt.figure(figsize = (12,8))
sns.countplot(data = most_liked_content_in_movies, x = 'year', hue = 'listed_in')
plt.legend(loc = 'upper left', frameon = True, ncol = 2)
plt.title('Content being added over the years')
plt.show()
```



```
In [99]: # visualization on top actors across tv shows in Netflix
plt.figure(figsize=(13,7))
plt.title('Top 10 Actor TV Shows Based on The Number of Titles and high rating')
sns.countplot(y = tv_show_data['cast'], order = tv_show_data['cast'].value_counts().index[:10], palette='pastel')
plt.show()
```

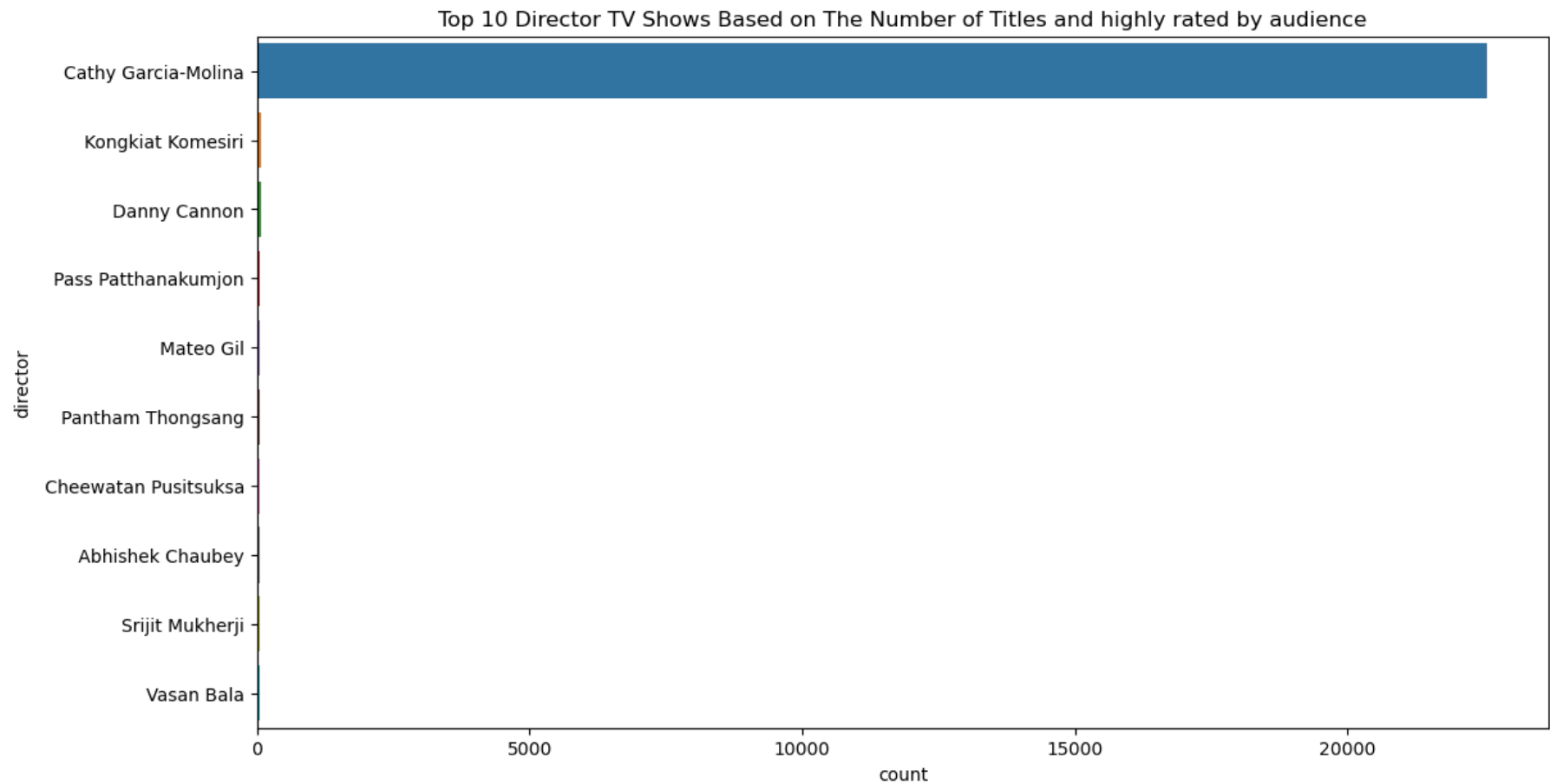


```
In [100]: # visualization on top actors across movies Netflix
plt.figure(figsize=(13,7))
plt.title('Top 10 Actor Movies Based on The Number of Titles and highly rated by audience')
sns.countplot(y = movies_data['cast'], order=movies_data['cast'].value_counts().index[:10], palette='pastel')
plt.show()
```

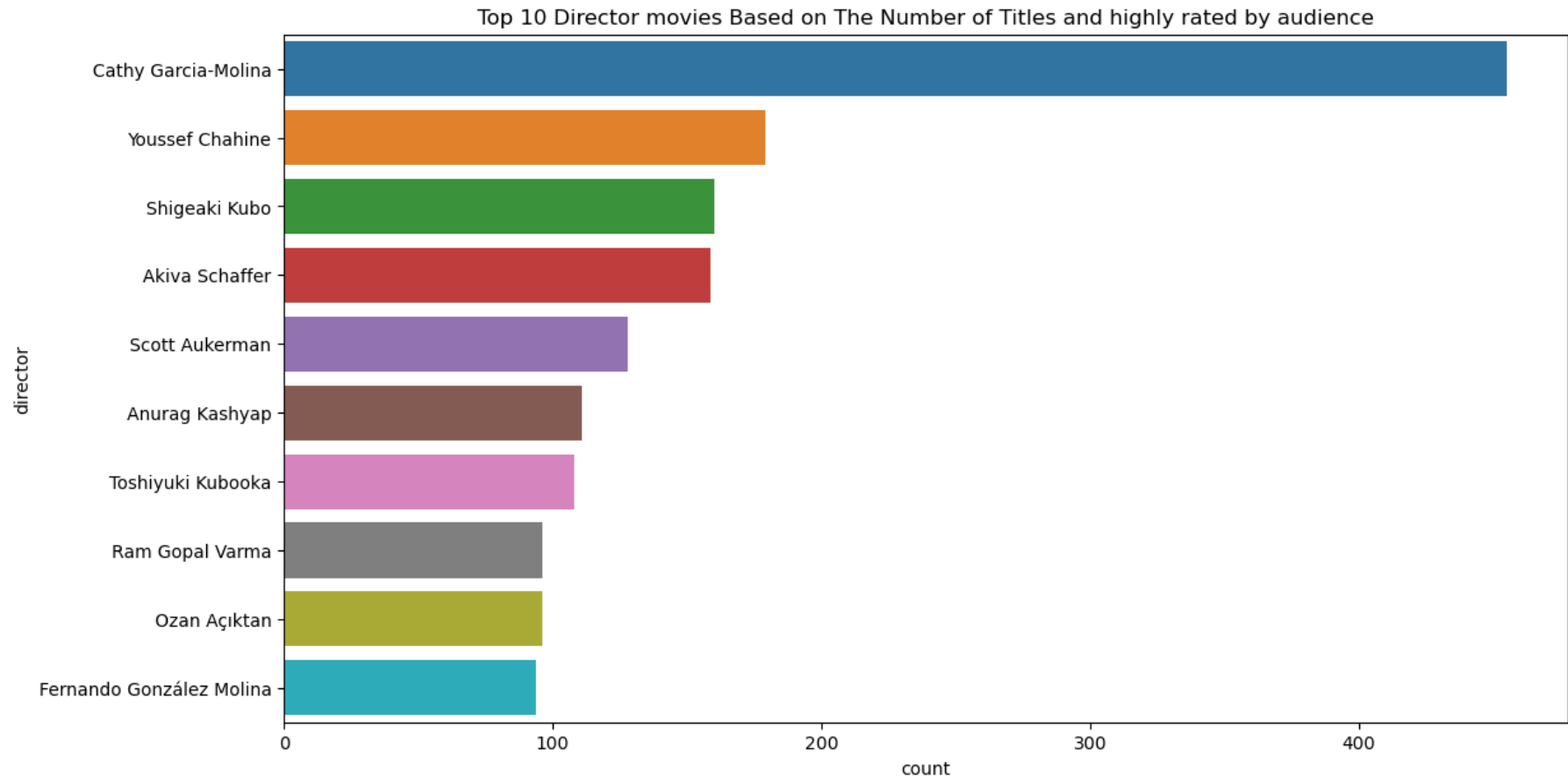




```
In [101]: # visualization on top directors across tv shows in Netflix
plt.figure(figsize=(13,7))
plt.title('Top 10 Director TV Shows Based on The Number of Titles and highly rated by audience')
sns.countplot(y = tv_show_data['director'], order=tv_show_data['director'].value_counts().index[:10])
plt.show()
```



```
In [102]: # visualization on top directors across movies in Netflix
plt.title('Top 10 Director movies Based on The Number of Titles and highly rated by audience')
sns.countplot(y = movies_data['director'], order=movies_data['director'].value_counts().index[:10])
plt.show()
```



```
In [103]: # countries producing the content that has been liked by everyone
countries_data = df.loc[df['rating'] == 'TV-MA', ['country', 'cast', 'director']]
```

```
In [104]: # which part of the world are generating a good rating
countries_data['country'].value_counts()
```

```
Out[104]: United States      15868
          India              5922
          Spain             3095
          United Kingdom     2929
          Japan             2828
          ...
          United States, China, United Kingdom      1
          United States, Australia                 1
          United States, United Kingdom, Germany    1
          Canada, India, Thailand, United States, United Arab Emirates 1
          United States, Denmark                   1
          Name: country, Length: 314, dtype: int64
```

```
In [105]: # highest rating recieved for these content with a certain duration being noted, irrespective of movies and shows
rating_data = pd.DataFrame(df.loc[df['rating'] == 'TV-MA',['country','duration','type','listed_in']])
```

In [106]: rating\_data

Out[106]:

	country	duration	type	listed_in
1	South Africa	2 Seasons	TV Show	International TV Shows
1	South Africa	2 Seasons	TV Show	TV Dramas
1	South Africa	2 Seasons	TV Show	TV Mysteries
1	South Africa	2 Seasons	TV Show	International TV Shows
1	South Africa	2 Seasons	TV Show	TV Dramas
...	...	...	...	...
8801	United Arab Emirates, Jordan	96 min	Movie	International Movies
8801	United Arab Emirates, Jordan	96 min	Movie	Thrillers
8801	United Arab Emirates, Jordan	96 min	Movie	Dramas
8801	United Arab Emirates, Jordan	96 min	Movie	International Movies
8801	United Arab Emirates, Jordan	96 min	Movie	Thrillers

61956 rows × 4 columns

In [307]: rating\_data['listed\_in'].value\_counts()

Out[307]:

International Movies	9216
Dramas	5553
TV Dramas	4494
Comedies	4068
International TV Shows	3585
...	
Classic & Cult TV	11
Classic Movies	10
Anime Features	10
Romantic Movies	10
Kids' TV	7

Name: listed\_in, Length: 67, dtype: int64

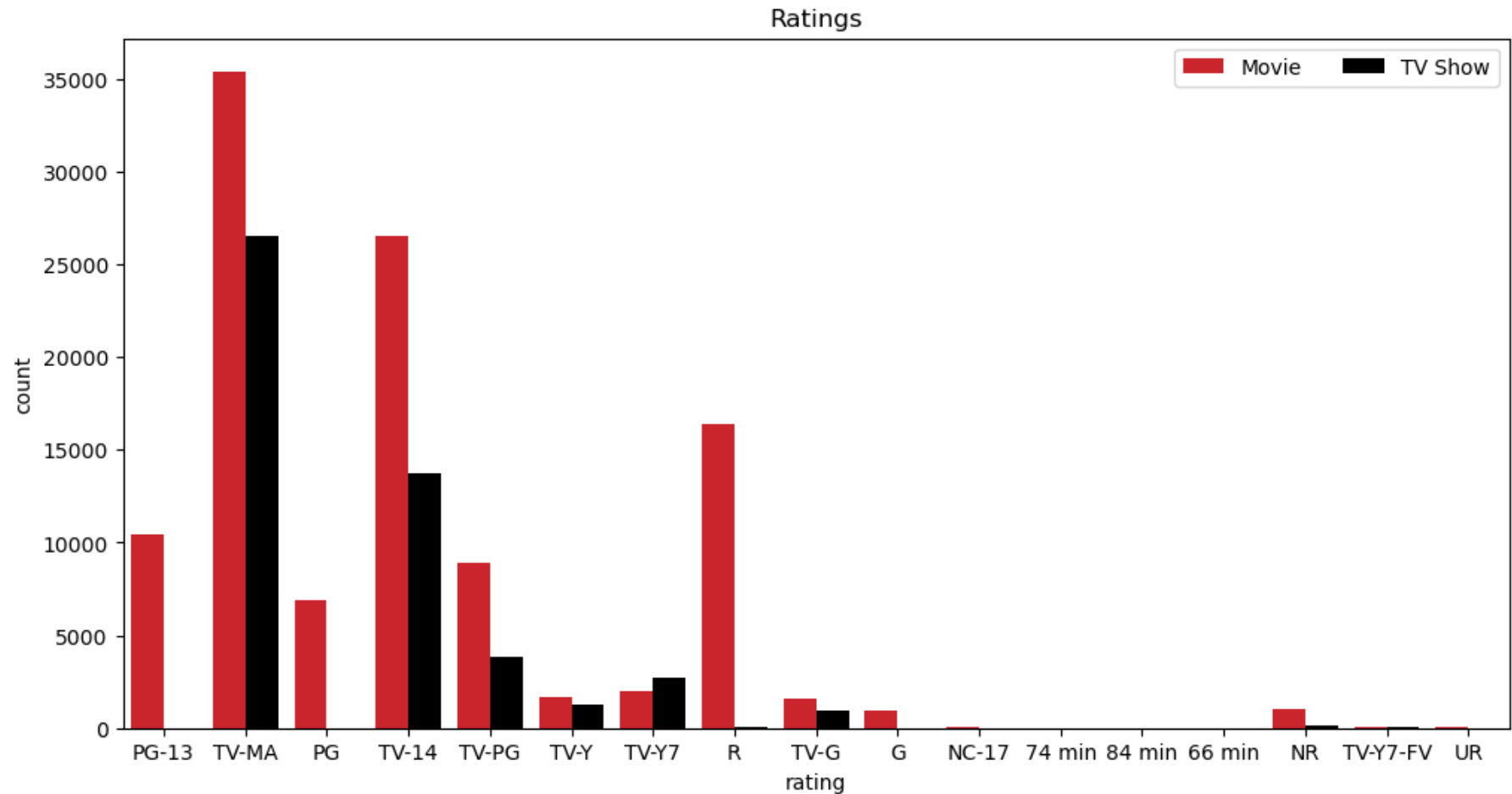
```
In [107]: rating_data['duration'].value_counts().head(20)
```

```
Out[107]: 1 Season      16916
          2 Seasons    4655
          3 Seasons    2536
          97 min       1146
          94 min       1041
          106 min      1036
          102 min       992
          99 min       940
          95 min       896
          107 min      890
          96 min       882
          105 min      871
          104 min      858
          103 min      857
          100 min      848
          91 min       794
          101 min      779
          4 Seasons    773
          5 Seasons    758
          92 min       743
          Name: duration, dtype: int64
```

```
In [108]: rating_data['type'].value_counts()
```

```
Out[108]: Movie      35410
          TV Show    26546
          Name: type, dtype: int64
```

```
In [131]: # checking the ratings for content being streamed in Netflix for both Movies and TV shows
plt.figure(figsize = (12,6))
sns.countplot(data = df, x = 'rating', hue = 'type',palette=['#E50914',"#000000"])
plt.legend(loc = 'upper right',frameon = True, ncol = 2)
plt.title('Ratings')
plt.show()
# Visualization show cases the the ratings based upon the movies and tv shows in Netflix
```



```
In [141]: # checking for least viewed genre based upon the least ratings received across tv shows and movies in Netflix  
least Rated genres = ['TV-Y', 'TV-Y7', 'NR', 'G', 'TV-Y7-FV', 'UR', 'NC-17', 'TV-G rating']
```

```
In [142]: least Rated content = df.loc[df['rating'].isin(least Rated genres), ['listed_in', 'cast', 'director']]
```

```
In [143]: leastRatedContent['listed_in'].value_counts()
```



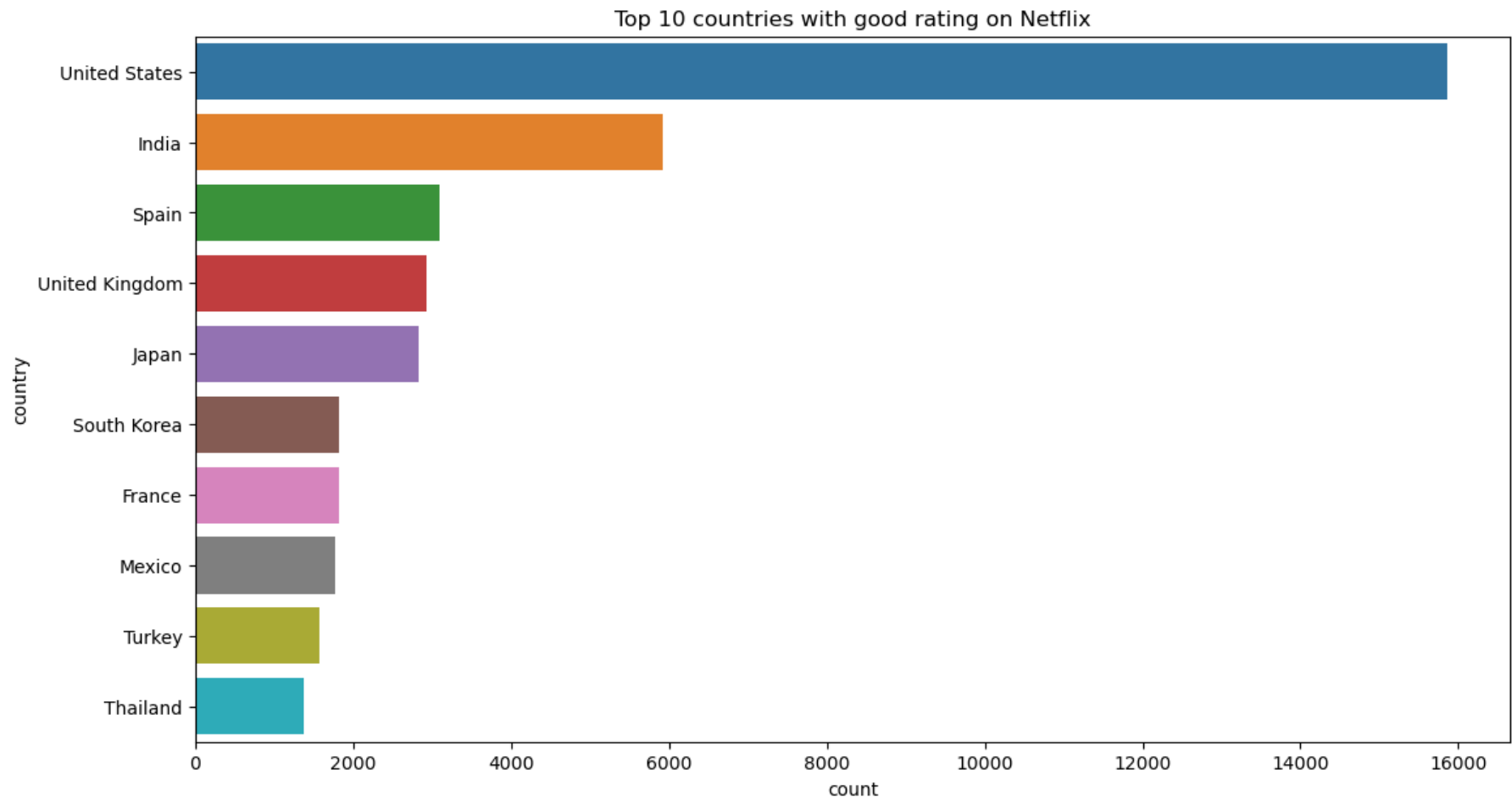
```

Out[143]: Children & Family Movies      1728
          Kids' TV                      828
          Comedies                      470
          International Movies          268
          Independent Movies           170
          Dramas                       168
          TV Comedies                   139
          Comedies                     136
          Dramas                       135
          British TV Shows             128
          Kids' TV                     114
          Music & Musicals              105
          Action & Adventure            99
          Movies                       93
          Classic Movies                76
          Classic Movies                58
          Korean TV Shows               56
          Documentaries                 46
          Romantic Movies               32
          Thrillers                     31
          Horror Movies                 30
          Sports Movies                 29
          Horror Movies                 26
          Spanish-Language TV Shows    25
          Romantic TV Shows             24
          Sci-Fi & Fantasy               22
          Cult Movies                   19
          TV Sci-Fi & Fantasy            19
          Crime TV Shows                17
          International TV Shows        17
          LGBTQ Movies                  15
          International TV Shows        14
          Anime Series                  13
          Stand-Up Comedy               12
          Cult Movies                   10
          TV Dramas                     10
          TV Action & Adventure          10
          TV Action & Adventure           8
          TV Thrillers                   6
          Classic & Cult TV              4
          Documentaries                  2

```

```
International Movies      1
Children & Family Movies 1
Docuseries               1
Stand-Up Comedy & Talk Shows 1
Name: listed_in, dtype: int64
```

```
In [110]: # top countries producing content in the netflix
plt.figure(figsize=(13,7))
plt.title('Top 10 countries with good rating on Netflix')
sns.countplot(y = rating_data['country'], order=rating_data['country'].value_counts().index[:10])
plt.show()
```



In [111]: *# best time to launch a new tv show in Netflix*

```
best_time_analysis = pd.DataFrame(df.groupby(['year', 'month']).aggregate(list_count = ('title', 'count')).reset_index())
```

In [112]: best\_time\_analysis

Out[112]:

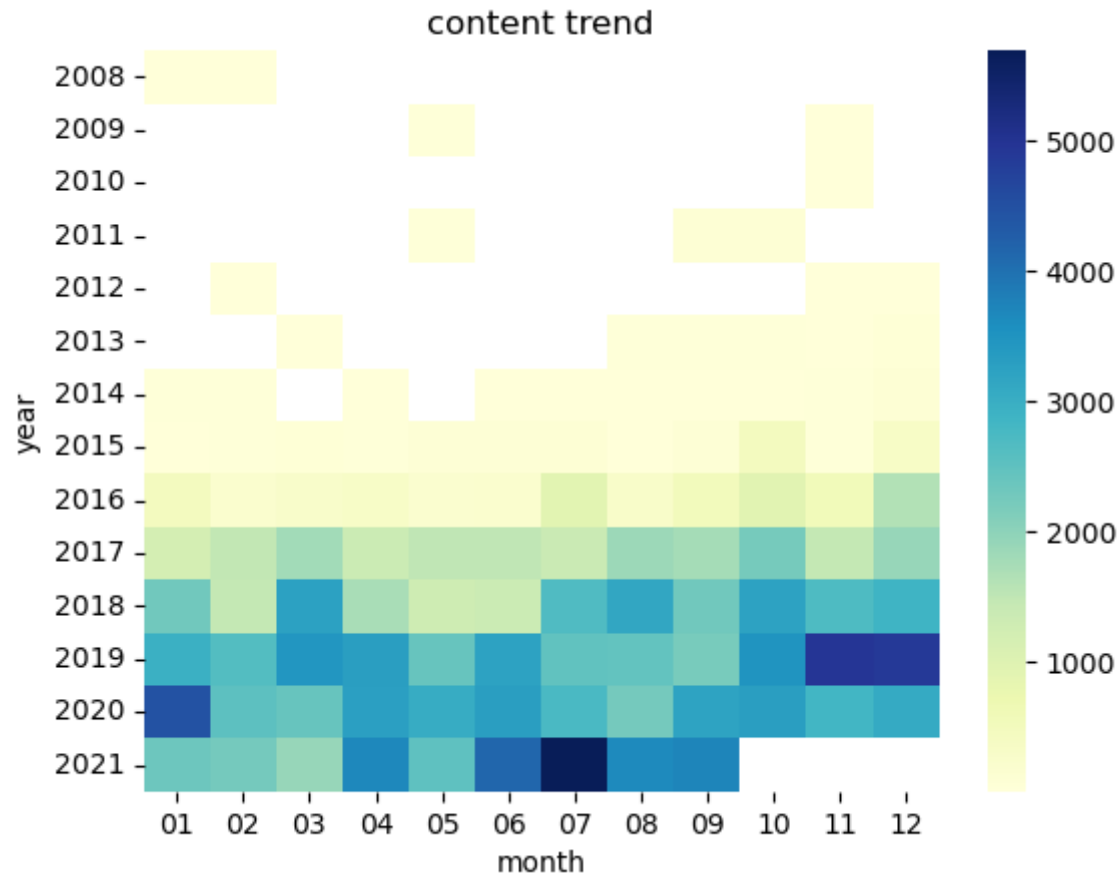
	year	month	list_count
0	2008	1	18
1	2008	2	1
2	2009	5	24
3	2009	11	6
4	2010	11	20
...	...	...	...
103	2021	5	2507
104	2021	6	4139
105	2021	7	5695
106	2021	8	3666
107	2021	9	3720

108 rows × 3 columns

In [155]: streaming\_data = best\_time\_analysis.pivot\_table(index = 'year', columns = 'month', values = 'list\_count')

In [166]: streaming\_data

```
In [130]: # visualizing the heatmap to analyze and predict the right time to launch the content in Netflix
sns.heatmap(streaming_data,cmap = 'YlGnBu',
            yticklabels = ['2008','2009','2010','2011','2012','2013','2014','2015','2016','2017','2018','2019','2020','2021'],
            xticklabels = ['01','02','03','04','05','06','07','08','09','10','11','12'])
plt.title('content trend')
plt.show()
```





```
In [ ]: # INSIGHTS & RECOMMENDATIONS
```

```
'''
```

```
Netflix EDA
```

```
Stats on movies trend in across years:
```

```
INSIGHTS:
```

- > From The amount of Content that has been generated in the Netflix, Movies has 69.5% of occupancy and 30.5% of occupancy with tv shows.
- > From the past 30 years there is significant increment in the number of movies being released till early 2019 but from then on the count of movies being released has been declining, can we say that COVID might be one of the key cause and can be several other factors dependent on it.
- > Precisely trend of movies released over the years has took a good incline graph but took a down fall from 2019.

```
-----  
Stats on Movies and Tv shows in Netflix:
```

```
INSIGHTS:
```

- > While we take a comparison on the count of tv shows and movies Netflix has been streaming in total:
- > The count of shows is pretty less compared to Movies, can say almost 50% or more lesser than movies.
- > We have a good quantity of movie content in Netlix.
- > But from the past couple of years Netflix has been investing on TV shows by which we can some decent amount of increment in tv shows being added into the netflix every year, As a supporting statement to that, in the recent years there are more number of shows being released compared to movies. And more over the amount of content added in the netflix supported tv shows while we can see a steady decrement of movies in the netflix, which took a step up and incremented the count of tv shows at the same time.
- > This might be because of the trend that in the covid times everyone are pretty much attracted to watching tv shows, while an evident proof supporting to that statement is, tv shows has gained a good amount of rating on the comparison with good rating between movies and tv shows. With a very less margin movies has better rating but still the kind of rating tv shows received in these recent years is commendable.
- > While we look out for ratings and the graph of ratings, TV MA ranks top, while least being NR

## RECOMMENDATIONS:

- > With only minimal content in the website and that too with limited number of tv shows, if the tv shows are gaining good amount of rating, it is better to launch them more in Netflix as the trend indicates that people are most likely tend to watch tv shows, which we can understand from the rating received by them in the recent years.
- > This can help Netflix generate a good amount of revenue from them and at the same time bring in a lot more good content as per the user preferrabilities.

-----

What is the best time to launch a TV show?

## INSIGHTS & RECOMMENDATIONS:

- > From the data being gathered and heatmap being generated, the best time to launch a time show is at post summer vacation at months of June, July or the other better time is at Nov, Dec where christmas vacation followed by a New year is other go to time, as we can observe there has been a huge amount of content being watched at these two scenarios.
- > Based on the heatmap being generated, we gathered this insights based upon the content that been added in those months more compared to other months. Amount of content generated high which is direct replica of content getting liked by the user watching the movies or tv shows which were released over that point of time in the year.

-----

Stats on actors and directors in Netflix:

## INSIGHTS:

- > Top 5 Directors based on number of tv shows titles in Netflix are Cathy Garcia-Molina, Rob Seidenglanz, Kongkiat Komesiri, Danny Cannon, Pass Patthanakumjon.
- > Top 5 Directors based on number of movies titles in Netflix are Cathy Garcia-Molina, Youssef chachine, Martin Scorsese, Dawid Dhawan, Steven Spielberg
- > And among the highly rated tv shows and movies, these are some insights on actors and directors:

- > Among the content that has been highly rated in the tv shows category, Anupam Kher, Hsia Teng-hung, Takahiro Sakurai, River Huang, Janel Tsai are the actors that are most liked and highly rated.
- > Among the content that has been highly rated in the tv shows category Cathy Garcia-Molina, Kongkiat Komesiri, Danny Cannon, Pass Patthanakumjon, Mateo Gil are directors whose tv shows are most liked ones and highly rated.
- > Among the content that been highly rated in tv shows these genres: TV Dramas, International TV Shows, Crime TV Shows, TV Comedies, Spanish-Language TV Shows, Romantic TV Shows are highly rated genres and users go to watch out ones.
- > Among the content that been highly rated in Movies these genres: International movies, Dramas, comedies, Independent movies, thrillers, action, romantic stand out as the highly rated by audience.
- > Among the content that has been highly rated in the Movies category: Anupam Kher, Radhika Apte, Piaa Bajpai, Usha Uthup, Rajat Kapoor are the actors that are most liked ones and part of highly rated movies.
- > But Anupam kher is the common actor both in tv shows and movies as the most liked and highly rated among there, can say almost beating other actors by a huge margin.
- > Among the content that has been highly rated in the movies category: Cathy Garcia-Molina, Youssef Chahine, Shigeaki Kubo, Akiva Schaffer, Scott Aukerman are directors whose movies are most liked ones and user go to movies can say.
- > And from asia the only director's movies that are highly rated are of Anurag Kashyap and Ram Gopal Varma.
- > And among the content that is highly rated, tv shows with duration 1 season, 2 season, 3 seasons are highly viewed and rated as the best.  
And movies with duration 97 min, 94, 106, 102, 99 are highly viewed and rated as best ones.
- > And among the content that has been least viewed and least rated: children & family movies, kids tv, british tv shows, classic movies, sports movies, music & musicals, documentaries, stand up comedy and some portion of international movies, dramas, independent movies, dramas, comedies are also least viewed and least rated.
- > Netflix needs to consider this as a recommendation to avoid acquiring this kind of content a bit linearly.

#### RECOMMENDATIONS:



- > Now with the insights that has been gained, we can strongly consider few things there and among them
- > Cathy Garcia-Molina movies are most liked and highly rated, due to which it gives us the opportunity to buy out the movies of him as as watch point of retain our users.
- > Simalarly the movies which stares Anupam Kher has gone mad in terms of viewing as the audience most likely went to watch his movies on a good percentage. this gives us the flexibility to get hold of the movies he is part.  
That really gives us a splendid opportunity to raise our user base to a good extent.
- > And while coming to the genres, our main focus should be on TV Dramas, International TV Shows, Crime TV Shows, TV Comedies, Spanish-Language TV Shows, Romantic TV Shows while buying the content in terms of tv shows.
- > Accordingly need to make sure that the movies with genres International movies, Dramas, comedies, Independent movies, thrillers, action, romantic should be given primary priority before buying the content.
- > Actors Anupam Kher, Hsia Teng-hung, Takahiro Sakurai, River Huang, Janel Tsai across tv shows should be considered as the go to films to buy in term where they are part of in tv shows category. And while coming to the movies category Anupam Kher, Radhika Apte, Piaa Bajpai , Usha Uthup , Rajat Kapoor movies should be considered as the priotiy picks analyzing their craze across the platform.
- > Cathy Garcia-Molina, Kongkiat Komesiri,Danny Cannon ,Pass Patthanakumjon, Mateo Gil directed tv shows across tv shows and Cathy Garcia-Molina, Youssef Chahine,Shigeaki Kubo ,Akiva Schaffer, Scott Aukerman movies across movies category should be the primary picks in buying.
- > User are watching the content with a specific critera of duration and among them 1 season, 2 season , 3 seasons are highly viewed across tv shows and 97 min, 94, 106,102,99 duration across movies should be considered. This gives the opportunity to buy out the movies or tv shows which are more likely to be in this category of durations.
- > As a note point, we need to avoid genres with children & family movies, kids tv, british tv shows, classic movies, sports movies, music & musicals, documentaries, stand up comedy and some portion of international movies, dramas, independent movies, dramas, comedies as they are pretty least viewed and heavily disliked by the audience. So avoiding such content can bring down our expenses and can put them to buy a more better content compared to it.  
This can also help us to hold the user trust and give the feeling that we are actually respecting their choice of watching the content and make Netflix as a user preferrable and worth content in their valuable time.

---

Stats on Netflix focusing on tv shows in recent years:

- > Can kind of say Yes and No to this scenario, if we take up them individually, yes the count of tv shows streamed has been constantly increasing in the past 5 years.
- > but while we compare this scenario with the count of Movies streamed in the past 10 years, the count of movies beat the tv shows with more than 50%, so this is not a right factor
- > Although Netflix has showed good interest in TV shows, still streaming the Movies is more than them.
- > yes they are focusing on tv shows in the recent years, there is steady increase in the content.

RECOMMENDATIONS:

- > Supporting the previous segment statements, Netflix has seen a good amount of rating being generated from tv shows and started investing on them. And as per the audience rating there are tending towards the tv shows more and this is the starting point to hold on with the content which is being liked by users and a good chance to make them stick to the application, producing on the tv show content with their suitable preferences among genres.
- > Encouraging this trend is only going to help Netflix build a solid platform of user retention and enhancement of pricing policy there by generating a good revenue up head.

---

Stats on content in different countries.

- > Top genres in Netflix according to the quantity are INTERNATIONAL MOVIES, DRAMAS, COMEDIES, ACTION & ADVENTURE, TV DRAMAS, INDEPENDENT MOVIES, CHILDREN & FAMILY MOVIES, INTERNATIONAL TV SHOWS, ROMANTIC MOVIES
- > Top countries producing content in netflix are US,INDIA JAPAN, UK, SOUTH KOREA, SPAIN.
- > Content in particular to Movies, the amount of movie content in Netflix is richly being added by countries United States standing first in the order following by India, Japan, UK and Nigeria in top 5.
- > while we look at the count of duration in movies, there are large number of movies with 94 minute duration, followed by 97,95,106, with least being 103.
- > Content in particular to tv shows, the amount of tv show content in Netflix is richly being added by countries United States standing first followed by Japan, south korea, UK and taiwan in top 5
- > while we look at the count of duration in tv shows, there are large number of tv shows with 1 season duration(highest) followed by 2,3,4, with least being 10 seasons.
- > considerably the duration in content is getting increased as the years are progressing for both tv shows and movies

RECOMMENDATIONS:

-> These are the stats related to the content being generated by different countries into the Netflix, the key point that can observe is United States ranked top among other in terms of producing content both in tv shows and movies. And to the duration part in each of the category movies with 94 min have more movies and tv shows with 1 season are more in Netflix.

-> If we want to draw a conclusion from this, from one of the visualization technique, we observed a point here that as the years are progressing, the duration of movies and tv shows both are increasing steadily in which we can make a conclusion that the audience are accepting the content even if it has more duration. And that gives us an insight of considering the factor that duration is not an issue, if at all we are watching or looking for a good content either in tv shows or movies.

...