GROUP1-CLASSWORK

October 20, 2022

1 WTF23 DATA SCIENCE AND ARTIFICAIL INTELLIGENCE¶

2 GROUP C SUBGROUP 1

3 CLASSWORK ON PYTHON- NUMPY, PANDAS AND MATPLOTLIP

[1]:
```
# importing the needed libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
netflix = pd.read_csv('netflix_titles.csv')
```

[3]:
```
# Dropping the show_id column
netflix = netflix.drop(columns=('show_id'))
netflix.head(5)
```

[3]:

	type	title	director	\
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	
1	TV Show	Blood & Water	NaN	
2	TV Show	Ganglands	Julien Leclercq	
3	TV Show	Jailbirds New Orleans	NaN	
4	TV Show	Kota Factory	NaN	
5	TV Show	Midnight Mass	Mike Flanagan	
6	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	
7	Movie	Sankofa	Haile Gerima	
8	TV Show	The Great British Baking Show	Andy Devonshire	
9	Movie	The Starling	Theodore Melfi	

	cast	\
0	NaN	
1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…	
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi…	
3	NaN	
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K…	

```
5  Kate Siegel, Zach Gilford, Hamish Linklater, H…
6  Vanessa Hudgens, Kimiko Glenn, James Marsden, …
7  Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D…
8  Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho…
9  Melissa McCarthy, Chris O'Dowd, Kevin Kline, T…

                                          country          date_added  \
0                                   United States  September 25, 2021
1                                    South Africa  September 24, 2021
2                                             NaN  September 24, 2021
3                                             NaN  September 24, 2021
4                                           India  September 24, 2021
5                                             NaN  September 24, 2021
6                                             NaN  September 24, 2021
7  United States, Ghana, Burkina Faso, United Kin…  September 24, 2021
8                                  United Kingdom  September 24, 2021
9                                   United States  September 24, 2021

   release_year rating   duration  \
0          2020  PG-13     90 min
1          2021  TV-MA  2 Seasons
2          2021  TV-MA   1 Season
3          2021  TV-MA   1 Season
4          2021  TV-MA  2 Seasons
5          2021  TV-MA   1 Season
6          2021     PG     91 min
7          1993  TV-MA    125 min
8          2021  TV-14  9 Seasons
9          2021  PG-13    104 min

                                       listed_in  \
0                                  Documentaries
1    International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act…
3                            Docuseries, Reality TV
4  International TV Shows, Romantic TV Shows, TV …
5                 TV Dramas, TV Horror, TV Mysteries
6                          Children & Family Movies
7   Dramas, Independent Movies, International Movies
8                       British TV Shows, Reality TV
9                                Comedies, Dramas

                                         description
0  As her father nears the end of his life, filmm…
1  After crossing paths at a party, a Cape Town t…
2  To protect his family from a powerful drug lor…
3  Feuds, flirtations and toilet talk go down amo…
```

```
4  In a city of coaching centers known to train I…
5  The arrival of a charismatic young priest brin…
6  Equestria's divided. But a bright-eyed hero be…
7  On a photo shoot in Ghana, an American model s…
8  A talented batch of amateur bakers face off in…
9  A woman adjusting to life after a loss contend…
```

[4]:
```
netflix.info() # exploring an overview of the data
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   type          8807 non-null   object
 1   title         8807 non-null   object
 2   director      6173 non-null   object
 3   cast          7982 non-null   object
 4   country       7976 non-null   object
 5   date_added    8797 non-null   object
 6   release_year  8807 non-null   int64 
 7   rating        8803 non-null   object
 8   duration      8804 non-null   object
 9   listed_in     8807 non-null   object
 10  description   8807 non-null   object
dtypes: int64(1), object(10)
memory usage: 757.0+ KB
```

[5]:
```
netflix.describe()
```

[5]:

	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

[6]:
```
netflix.isnull().sum() # examining the missing values in the dataset
```

[6]:
```
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
```

[7]:
```
# dropping the few null rows
netflix = netflix.dropna(subset=['date_added', 'rating', 'duration'])

# filling the null rows in the director column
netflix['director'] = netflix['director'].fillna('None Specified')

netflix['cast'] = netflix['cast'].fillna('Not Listed')

netflix['country'] = netflix['country'].fillna('Not Specified')
```

[8]:
```
# Ensuring there are no missing values
netflix.isnull().sum()
```

[8]:
```
type            0
title           0
director        0
cast            0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
description     0
dtype: int64
```

[9]:
```
# Converting the duration to an interger
netflix['duration'] = netflix['duration'].str.strip('min')
netflix['duration'] = netflix['duration'].str.strip('Seasons')
netflix['duration'] = netflix['duration'].str.strip('Seaso')
```

[10]:
```
# removing the extra whiite spaces and converting to an integer
netflix['duration'] = netflix['duration'].str.strip(' ').astype('int')
```

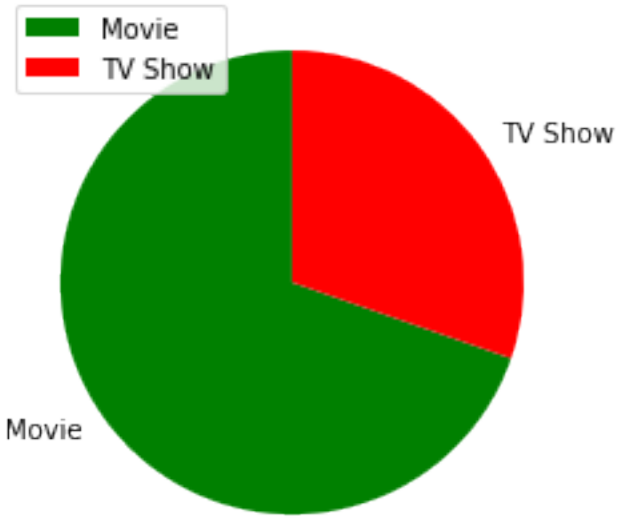
[11]:
```
netflix['type'].value_counts()
```

[11]:
```
Movie      6126
TV Show    2664
Name: type, dtype: int64
```

3.1 Distribution of Movies and Series

This pie chart shows that there are a lot more movies produced than tv shows

[13]:
```
plt.pie(netflix['type'].value_counts(),
        labels=netflix['type'].value_counts().index,
        colors=['green', 'red'], startangle=90

       )
plt.legend();
```



3.2 Movie subset

[14]:
```
# subset the data to only include movies
netflix_movie = netflix[netflix['type'] == 'Movie']
netflix_movie.head()
```

[14]:

	type	title	director	\
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	
6	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	
7	Movie	Sankofa	Haile Gerima	
9	Movie	The Starling	Theodore Melfi	
12	Movie	Je Suis Karl	Christian Schwochow	

	cast	\
0	Not Listed	
6	Vanessa Hudgens, Kimiko Glenn, James Marsden, …	
7	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D…	

```
9   Melissa McCarthy, Chris O'Dowd, Kevin Kline, T…
12  Luna Wedler, Jannis Niewöhner, Milan Peschel, …

                                                country          date_added  \
0                                         United States  September 25, 2021
6                                         Not Specified  September 24, 2021
7   United States, Ghana, Burkina Faso, United Kin…  September 24, 2021
9                                         United States  September 24, 2021
12                              Germany, Czech Republic  September 23, 2021

    release_year rating  duration  \
0           2020  PG-13        90
6           2021     PG        91
7           1993  TV-MA       125
9           2021  PG-13       104
12          2021  TV-MA       127

                                           listed_in  \
0                                      Documentaries
6                              Children & Family Movies
7   Dramas, Independent Movies, International Movies
9                                   Comedies, Dramas
12                        Dramas, International Movies

                                          description
0   As her father nears the end of his life, filmm…
6   Equestria's divided. But a bright-eyed hero be…
7   On a photo shoot in Ghana, an American model s…
9   A woman adjusting to life after a loss contend…
12  After most of her family is murdered in a terr…
```

3.3 Top directors

[15]:
```
# getting the counts of directors in the dataset
netflix['director'].value_counts()
```

[15]:
```
None Specified                  2621
Rajiv Chilaka                     19
Raúl Campos, Jan Suter            18
Suhas Kadav                       16
Marcus Raboy                      16
                                …
Raymie Muzquiz, Stu Livingston     1
Joe Menendez                       1
Eric Bross                         1
Will Eisenberg                     1
Mozez Singh                        1
```

```
Name: director, Length: 4527, dtype: int64
```

[16]:
```
# Split the movie directors to only include 1 name
movie_directors = netflix_movie['director'].str.split(',', expand=True).stack()
# convert to a dataframe
movie_directors = pd.DataFrame(movie_directors)
movie_directors.head() # view the first few rows
```

[16]:

		0
0	0	Kirsten Johnson
6	0	Robert Cullen
	1	José Luis Ucha
7	0	Haile Gerima
9	0	Theodore Melfi

[17]:
```
# name the column
movie_directors.columns = ['Directors']
```

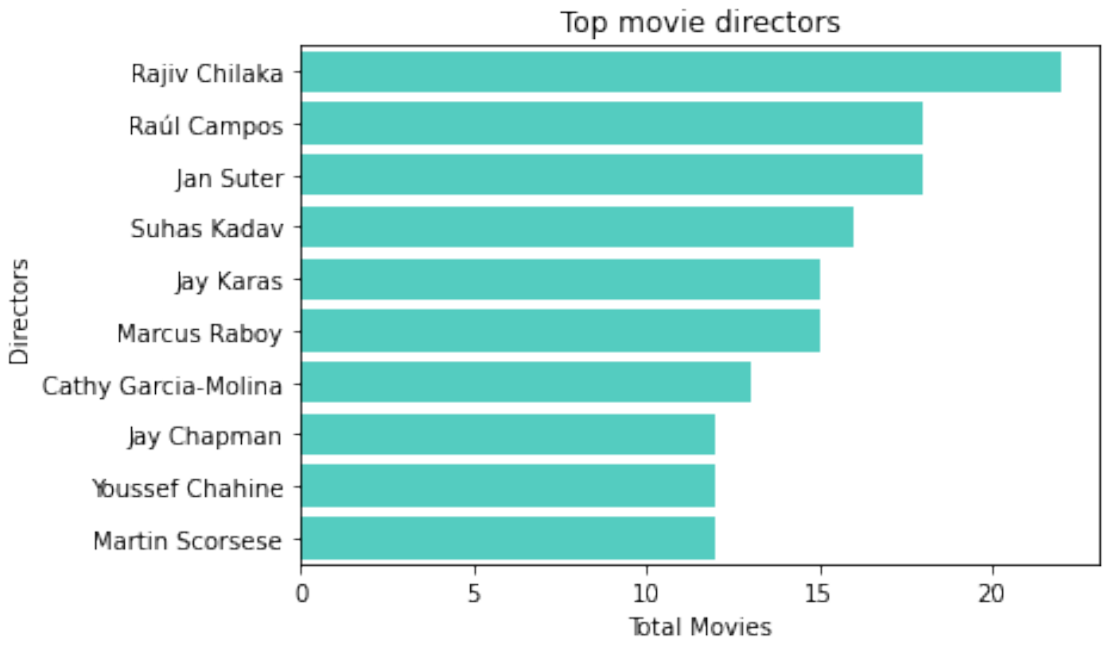
[18]:
```
# only selecting the known directors
movie_directors = movie_directors[movie_directors['Directors'] != 'None␣
 ↪Specified']

movie_directors = movie_directors.groupby(['Directors']).size().
 ↪reset_index(name='Total Movies')
```

[19]:
```
top_directors = pd.DataFrame(movie_directors.sort_values(by=['Total Movies'],␣
 ↪ascending=False)).head(10)
top_directors
```

[19]:

	Directors	Total Movies
3816	Rajiv Chilaka	22
3863	Raúl Campos	18
234	Jan Suter	18
4427	Suhas Kadav	16
2307	Jay Karas	15
3065	Marcus Raboy	15
1288	Cathy Garcia-Molina	13
2304	Jay Chapman	12
4838	Youssef Chahine	12
3133	Martin Scorsese	12

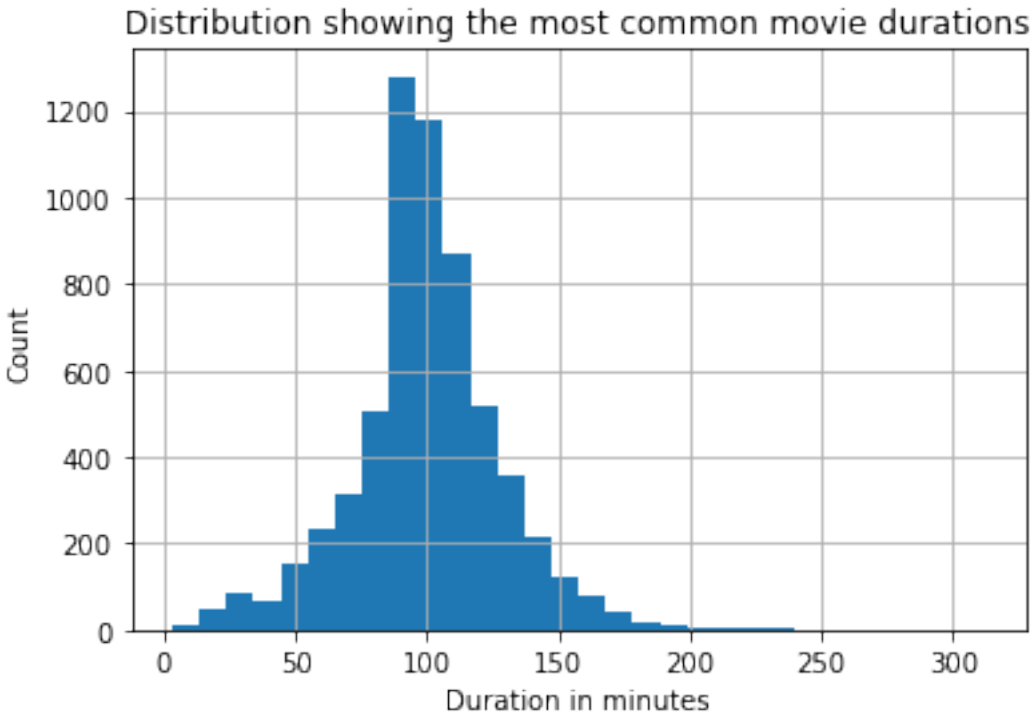
[20]:
```
# Visualising the top movie directors
sns.barplot(y='Directors', x='Total Movies', data=top_directors,␣
 ↪color='turquoise')
plt.title('Top movie directors');
```



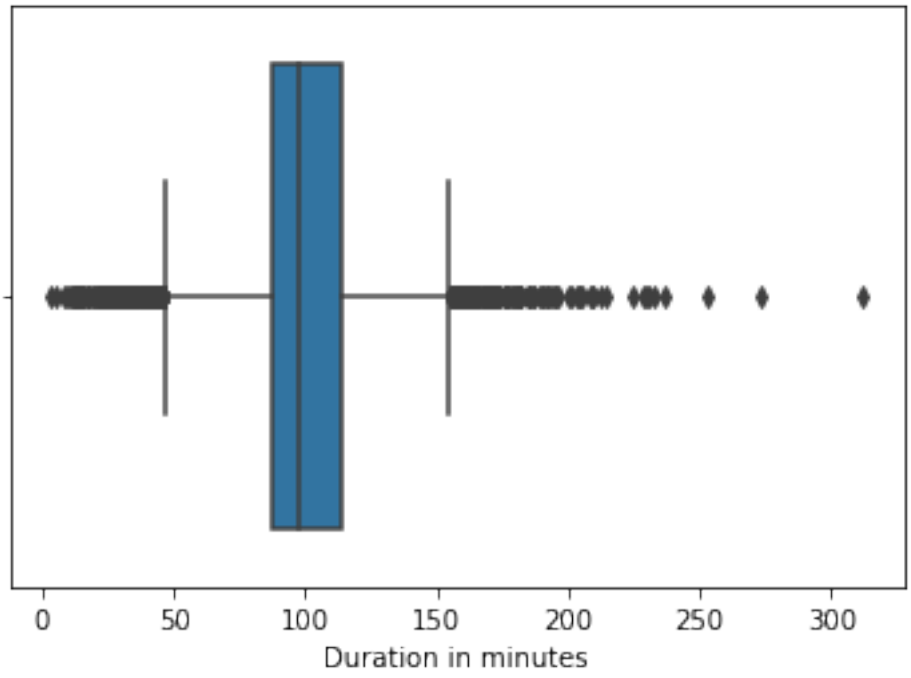
[21]:
```
# This shows that the top movie director is Rajiv Chilaka
```

3.4 Movie durations distribution

[22]:
```
netflix_movie['duration'].hist(bins=30)
plt.xlabel('Duration in minutes')
plt.ylabel('Count')
plt.title('Distribution showing the most common movie durations');
```



[23]:
```
# to show the outliers in the durations column
sns.boxplot(x='duration', data=netflix_movie)
plt.xlabel('Duration in minutes');
```



[24]:
```
#These plots show that most movies are between 80 to 100 minutes long
```

3.5 Top movie countries

[25]:
```
# Expanding the countries to only include one country
countries = netflix_movie['country'].str.split(',', expand=True).stack()
# removing the white sapces before and after country names to avoid duplicates
countries = countries.str.strip(' ')
```

[26]:
```
countries.value_counts()
```

[26]:
```
United States     2749
India              962
United Kingdom     534
Not Specified      439
Canada             319
                  …
Bermuda              1
Ecuador              1
Armenia              1
Mongolia             1
Montenegro           1
Length: 119, dtype: int64
```

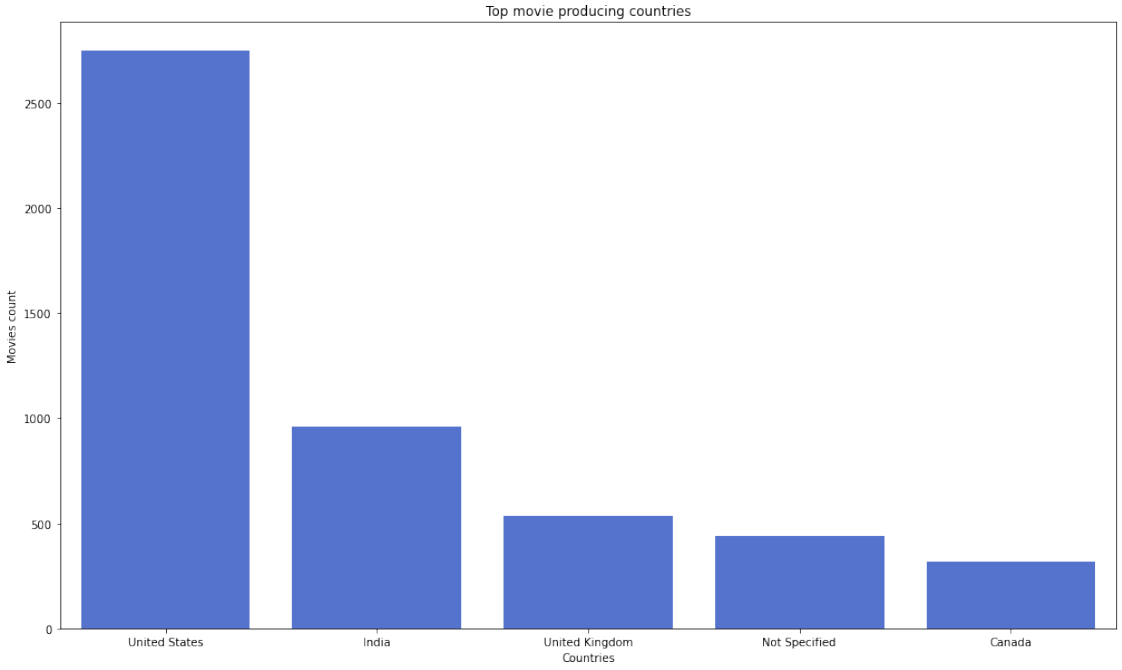
[27]:
```
top_countries = countries.value_counts().head(5)
```

[28]:
```
top_countries = pd.DataFrame(top_countries)
# renaming the column
top_countries.columns = ['Movies count']
top_countries
```

[28]:

	Movies count
United States	2749
India	962
United Kingdom	534
Not Specified	439
Canada	319

[29]:
```
# A bar plot to show the top movie producing countries
plt.figure(figsize=(17, 10))
sns.barplot(x=top_countries.index, y=top_countries['Movies count'],␣
 ↪data=top_countries, color='royalblue')
plt.xlabel('Countries')
plt.title('Top movie producing countries');
```



The above bar chart shows that United States is the largest movie producing country

3.6 Top movie actors

[30]:
```
actors = netflix_movie['cast'].str.split(',', expand=True).stack()
actors = pd.DataFrame(actors)
actors.head()
```

[30]:
```
                   0
0 0       Not Listed
6 0  Vanessa Hudgens
  1     Kimiko Glenn
  2    James Marsden
  3     Sofia Carson
```

[31]:
```
actors.columns = ['cast']
```

[33]:
```
actors = actors[actors['cast'] != 'Not Listed']
actors.head()
```

[33]:
```
                cast
6 0  Vanessa Hudgens
  1     Kimiko Glenn
  2    James Marsden
  3     Sofia Carson
```

```
4       Liza Koshy
```

3.7 Release year

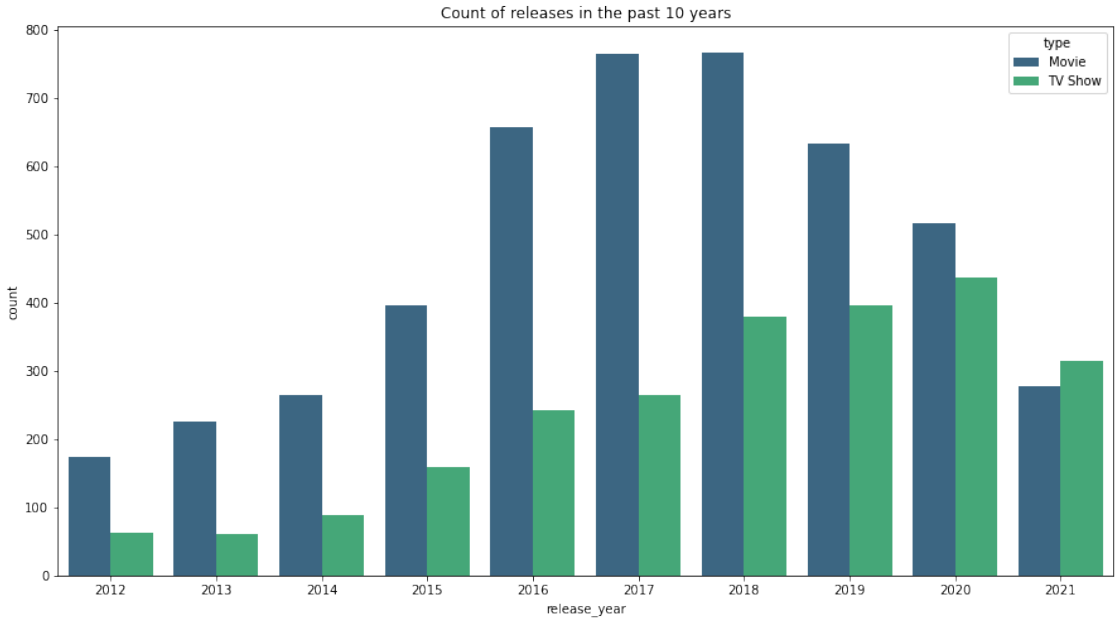
[40]:
```
netflix['release_year'].value_counts()
```

[40]:
```
2018    1146
2017    1030
2019    1030
2020     953
2016     901
        …
1959       1
1925       1
1961       1
1947       1
1966       1
Name: release_year, Length: 74, dtype: int64
```

[41]:
```
# Selecting two columns to perform a bi-variate analysis
netflix_release = netflix[['type','release_year']]
# Selecting the last 10 years
last_10_years = netflix_release[netflix['release_year'] >= 2012]
last_10_years
```

[41]:
```
         type  release_year
0       Movie          2020
1     TV Show          2021
2     TV Show          2021
3     TV Show          2021
4     TV Show          2021
…       …            …
8798    Movie          2014
8800  TV Show          2012
8801    Movie          2015
8803  TV Show          2018
8806    Movie          2015

[7081 rows x 2 columns]
```

[42]:
```
# A bar plot showing the counts of movies and series produced in the last 10␣
 ↪years
plt.figure(figsize=(15, 8))
sns.countplot(x='release_year', data=last_10_years, hue='type',␣
 ↪palette='viridis')
plt.title('Count of releases in the past 10 years');
```

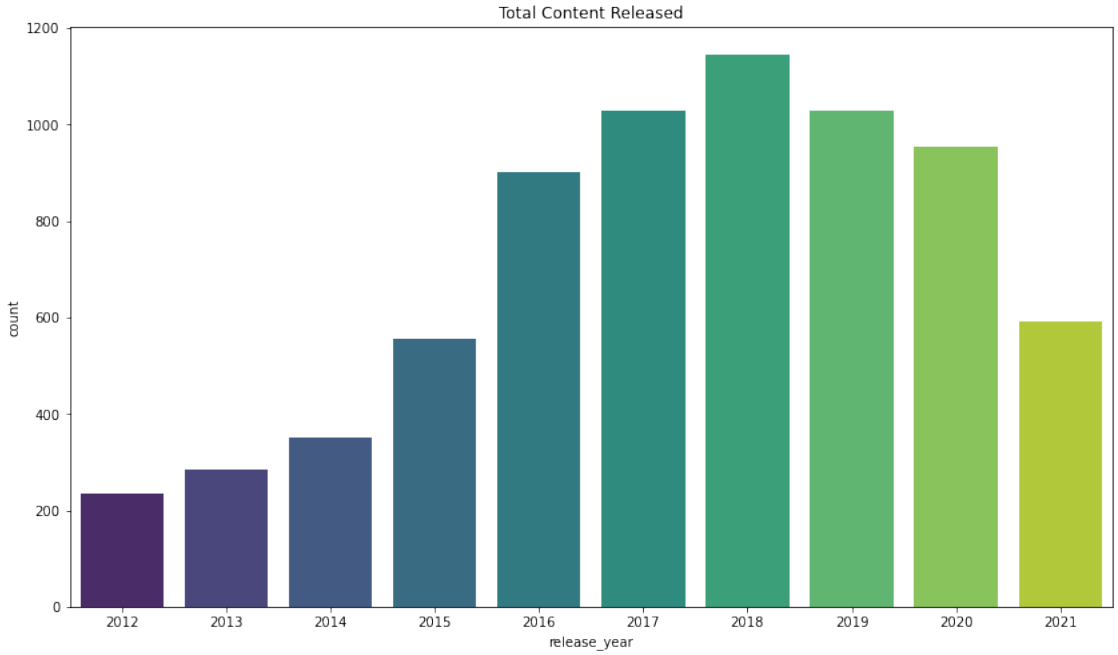


this bar plot hows that the highest number of movies produced were in the years 2017 and 2018.

It also shows that the highest number of tv shows were produced in 2020

**** COUNT PLOT OF TOTAL MOVIES AND TV SHOWS****

[44]:
```
# A count plot showing the total movies and Tv shows released
plt.figure(figsize=(14, 8))
sns.countplot(x='release_year', data=last_10_years, palette='viridis')
plt.title('Total Content Released');
```



The above count plot shows that the year 2018 had the most content produced

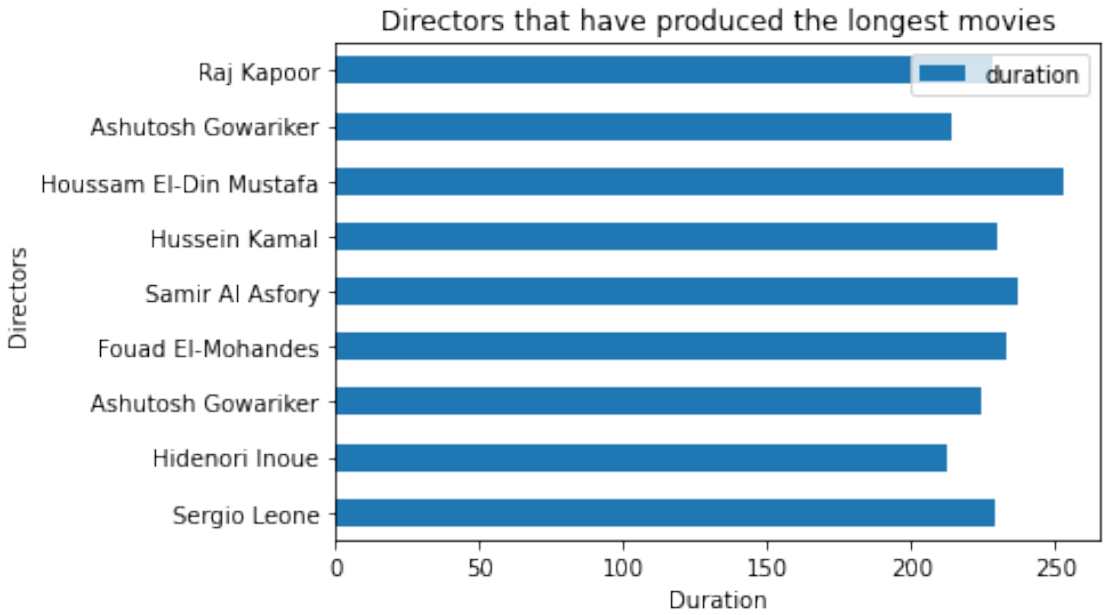
3.8 Directors with the longest movies

[45]:
```
longest_movies = netflix_movie[['director','duration']]
longest_movies= longest_movies[longest_movies['director'] != 'None Specified']
```

[46]:
```
directors_longest = longest_movies[longest_movies['duration'] > 210]
```

[47]:
```
directors_longest = directors_longest.set_index(['director'])
```

[48]:
```
directors_longest.plot(kind='barh')
plt.xlabel('Duration')
plt.ylabel('Directors')
plt.title('Directors that have produced the longest movies');
```



This plot shows the directors that have produced the longest movies is Houssam El-Din Mustafa

Visualization of title Ratings

[50]:
```
netflix['rating'].unique()
```

[50]:
```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
       'TV-G', 'G', 'NC-17', 'NR', 'TV-Y7-FV', 'UR'], dtype=object)
```

[54]:
```
netflix['rating'].value_counts()
```

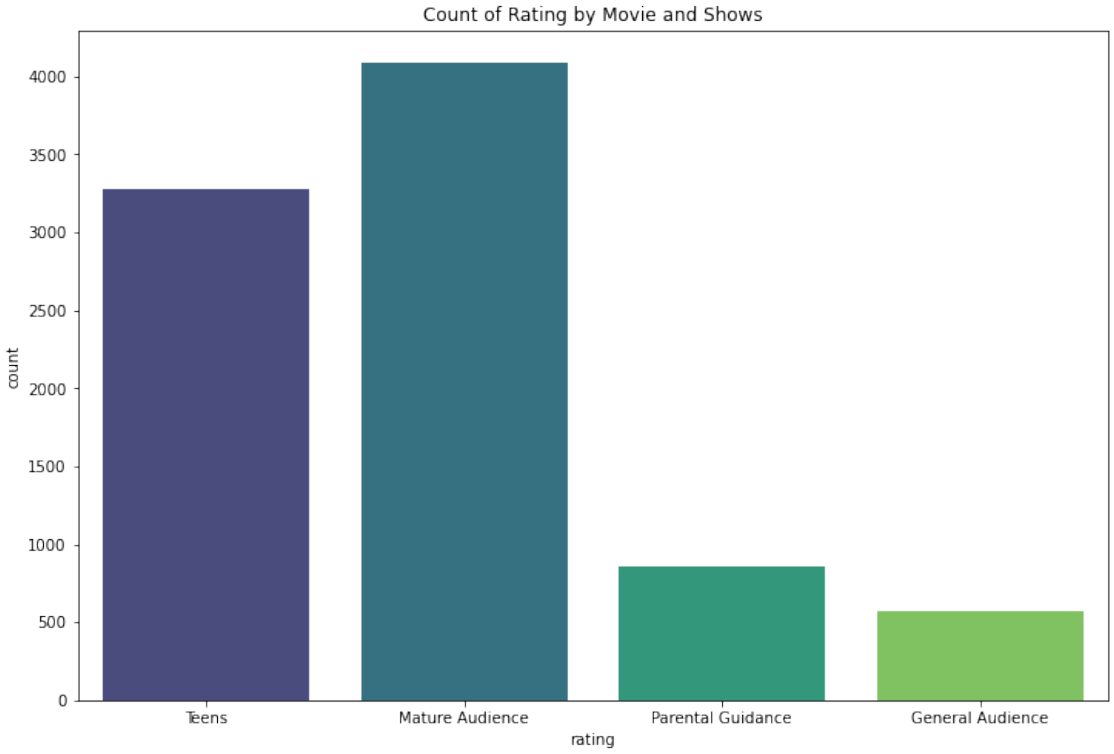
[54]:
```
Mature Audience      4089
Teens                3273
Parental Guidance     861
General Audience      567
Name: rating, dtype: int64
```

[55]:
```
# creating a dictionary to reorganise the ratings
new_categories = {
    'TV-PG': 'Parental Guidance',
    'TV-MA': 'Mature Audience',
    'TV-Y7-FV': 'Teens',
    'TV-Y7': 'Teens',
    'TV-14': 'Teens',
    'R': 'Mature Audience',
    'TV-Y': 'General Audience',
```

```
    'NR': 'Mature Audience',
    'PG-13': 'Teens',
    'TV-G': 'General Audience',
    'PG': 'Teens',
    'G': 'General Audience',
    'UR': 'Mature Audience',
    'NC-17': 'Mature Audience'
}
netflix["rating"] = netflix['rating'].replace(new_categories)
netflix['rating'].value_counts()
```

[55]:
```
Mature Audience      4089
Teens                3273
Parental Guidance     861
General Audience      567
Name: rating, dtype: int64
```

[56]:
```
plt.figure(figsize=(12,8))
sns.countplot(x="rating", data=netflix, palette="viridis")
plt.title("Count of Rating by Movie and Shows");
```



The visualization shows that most of the content released are for mature audiences

3.9 Visualization of Longest movie ratings

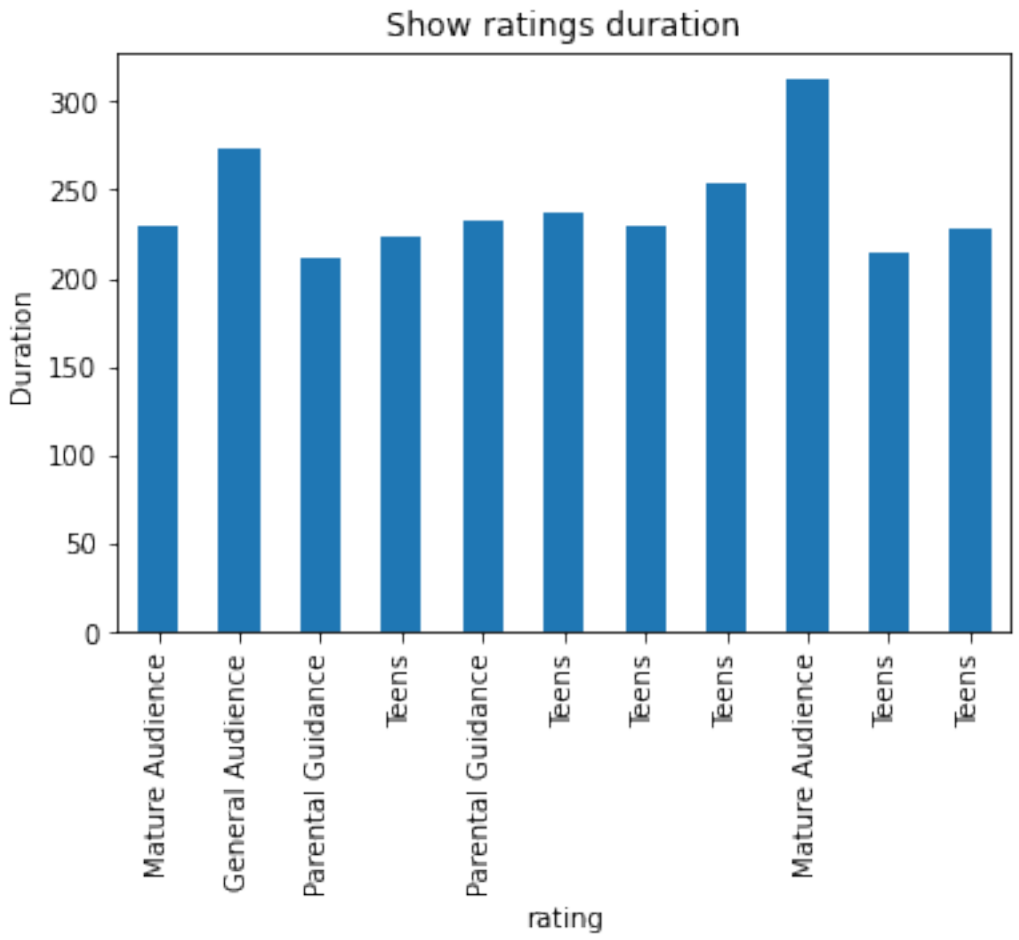
[58]:
```
ratings = netflix[['rating', 'duration']]
ratings.value_counts()
```

[58]:
```
rating             duration
Mature Audience    1           777
Teens              1           628
Parental Guidance  1           239
Mature Audience    2           190
General Audience   1           147
                               …
Mature Audience    312           1
                   229           1
                   209           1
                   208           1
Teens              253           1
Length: 610, dtype: int64
```

[59]:
```
# subsetting to only include the ratings with more than 210 minutes
longest_ratings = ratings[ratings['duration'] > 210]
# setting the index to the rating column
longest_ratings = longest_ratings.set_index(['rating'])
longest_ratings
```

[59]:
```
                   duration
rating
Mature Audience         229
General Audience        273
Parental Guidance       212
Teens                   224
Parental Guidance       233
Teens                   237
Teens                   230
Teens                   253
Mature Audience         312
Teens                   214
Teens                   228
```

[60]:
```
longest_ratings['duration'].plot(kind='bar')
plt.ylabel('Duration')
plt.title('Show ratings duration');
```



This shows that mature audience rating has the longest movies

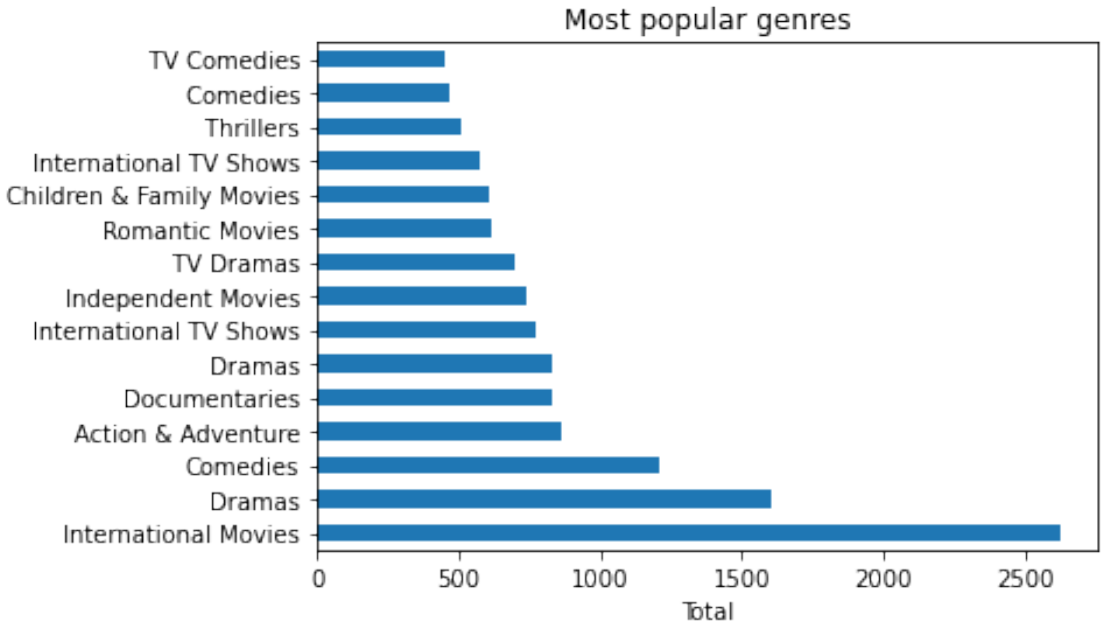
3.10 Most popular genres

[61]:
```
# converting the column to a dataframe
#netflix['listed_in']  = pd.DataFrame(netflix['listed_in'])
# Splitting each genre to allow for accurate counting
genres = netflix['listed_in'].str.split(',', expand=True).stack()
# Sorting the values and picking the top 15
popular = genres.value_counts().sort_values(ascending=False).iloc[:15]
popular
```

[61]:
```
 International Movies        2624
Dramas                       1599
Comedies                     1210
Action & Adventure            859
Documentaries                 829
```

```
 Dramas                      827
International TV Shows       773
 Independent Movies          736
 TV Dramas                   695
 Romantic Movies             613
Children & Family Movies     605
 International TV Shows      576
 Thrillers                   512
 Comedies                    464
 TV Comedies                 454
dtype: int64
```

[62]:
```
popular.plot(kind='barh')
plt.xlabel('Total')
plt.title('Most popular genres');
```



This shows that the genre international movies is the most popular

[63]:
```
netflix.head()
```

[63]:

	type	title	director	\
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	
1	TV Show	Blood & Water	None Specified	
2	TV Show	Ganglands	Julien Leclercq	
3	TV Show	Jailbirds New Orleans	None Specified	
4	TV Show	Kota Factory	None Specified	

```
                                                cast        country  \
0                                          Not Listed  United States
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…   South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi…  Not Specified
3                                          Not Listed  Not Specified
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K…          India

          date_added  release_year          rating  duration  \
0  September 25, 2021          2020           Teens        90
1  September 24, 2021          2021  Mature Audience         2
2  September 24, 2021          2021  Mature Audience         1
3  September 24, 2021          2021  Mature Audience         1
4  September 24, 2021          2021  Mature Audience         2

                                      listed_in  \
0                                  Documentaries
1    International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act…
3                            Docuseries, Reality TV
4  International TV Shows, Romantic TV Shows, TV …

                                         description
0  As her father nears the end of his life, filmm…
1  After crossing paths at a party, a Cape Town t…
2  To protect his family from a powerful drug lor…
3  Feuds, flirtations and toilet talk go down amo…
4  In a city of coaching centers known to train I…
```

[64]:
```
netflix['listed_in'].value_counts()
```

[64]:
```
Dramas, International Movies                                362
Documentaries                                               359
Stand-Up Comedy                                             334
Comedies, Dramas, International Movies                      274
Dramas, Independent Movies, International Movies            252
                                                           …
Crime TV Shows, International TV Shows, TV Sci-Fi & Fantasy   1
International TV Shows, TV Horror, TV Sci-Fi & Fantasy        1
Crime TV Shows, Kids' TV                                      1
Horror Movies, International Movies, Sci-Fi & Fantasy         1
Cult Movies, Dramas, Thrillers                                1
Name: listed_in, Length: 513, dtype: int64
```

[65]:
```
netflix_movie['listed_in'].value_counts()
```

[65]:
```
Dramas, International Movies                362
Documentaries                               359
```

```
Stand-Up Comedy                                        334
Comedies, Dramas, International Movies                 274
Dramas, Independent Movies, International Movies       252
                                                       …
Sci-Fi & Fantasy                                         1
Sports Movies                                            1
Children & Family Movies, Comedies, Cult Movies          1
Cult Movies, Dramas, Music & Musicals                    1
Cult Movies, Dramas, Thrillers                           1
Name: listed_in, Length: 278, dtype: int64
```

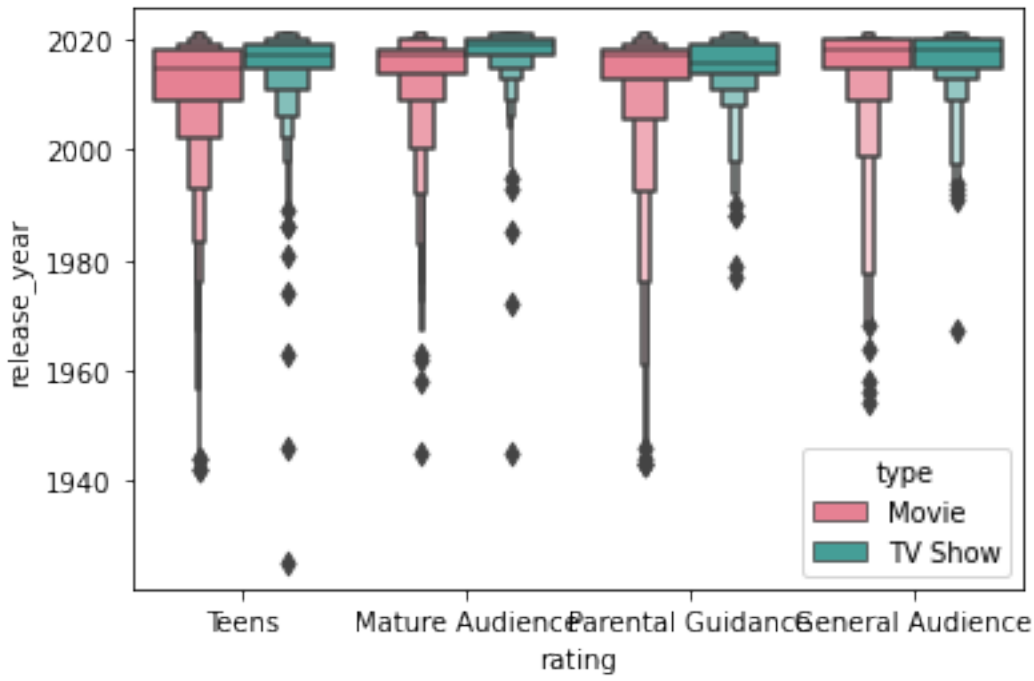
[66]:
```
top_ratings = netflix[['title','rating']]
top_ratings
```

[66]:
```
                      title           rating
0       Dick Johnson Is Dead             Teens
1               Blood & Water  Mature Audience
2                  Ganglands  Mature Audience
3      Jailbirds New Orleans  Mature Audience
4                Kota Factory  Mature Audience
…                        …               …
8802                  Zodiac  Mature Audience
8803             Zombie Dumb             Teens
8804              Zombieland  Mature Audience
8805                    Zoom             Teens
8806                  Zubaan             Teens

[8790 rows x 2 columns]
```

** MULTIVARIATE VISUALIZATION OF THE NETFLIX DATA**

[71]:
```
sns.boxenplot(x="rating", y="release_year", hue="type", data=netflix,␣
↪palette="husl")
```

[71]:
```
<AxesSubplot:xlabel='rating', ylabel='release_year'>
```



VISUALIZING THE DISTRIBUTION OF MOVIE RATING

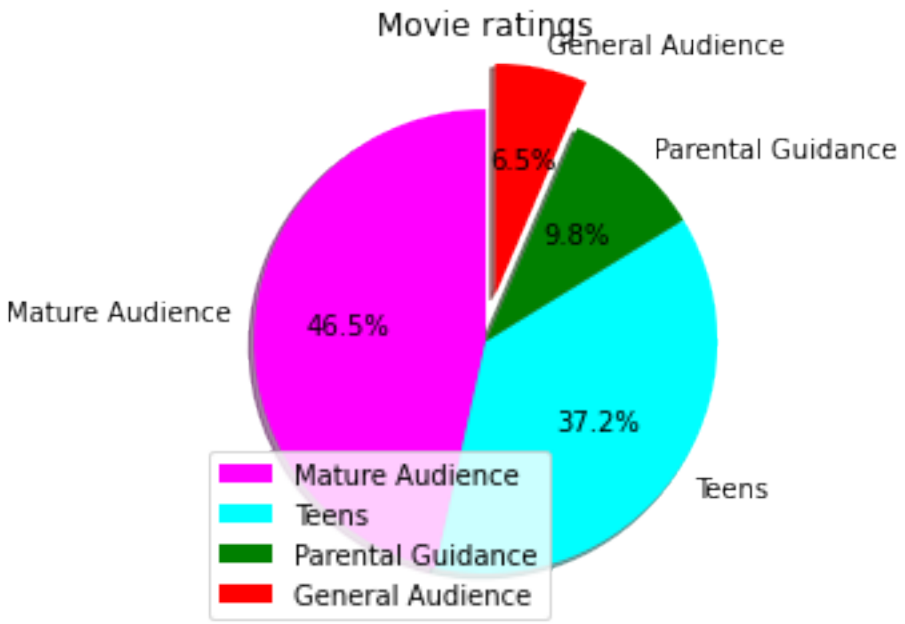
[75]:
```
ratings = top_ratings['rating'].value_counts()
ratings
```

[75]:
```
Mature Audience      4089
Teens                3273
Parental Guidance     861
General Audience      567
Name: rating, dtype: int64
```

[76]:
```
ratings.index
```

[76]:
```
Index(['Mature Audience', 'Teens', 'Parental Guidance', 'General Audience'],
dtype='object')
```

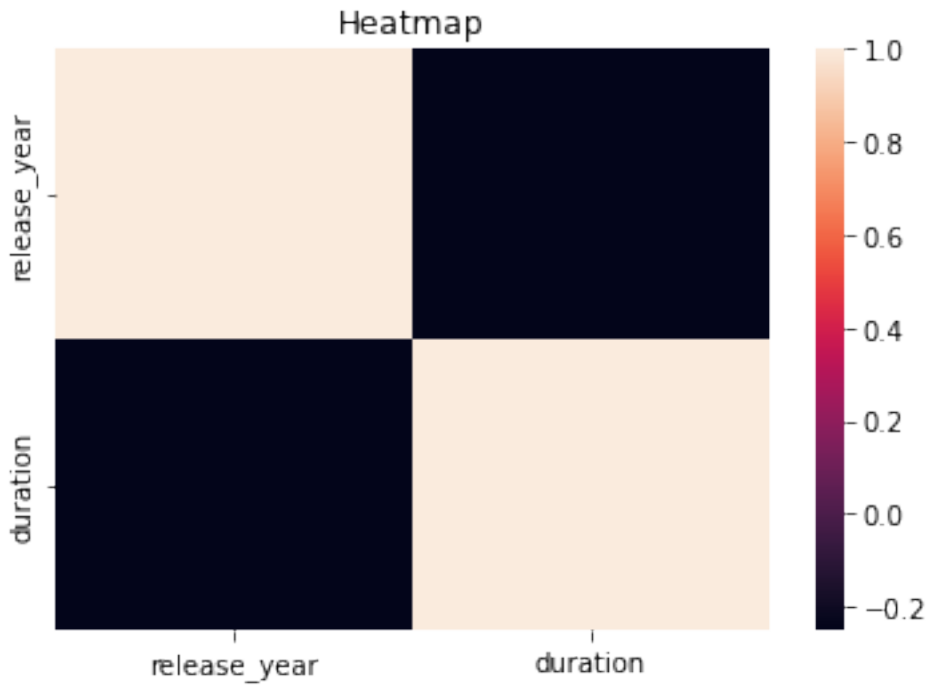
[77]:
```
# Displays the distribution of the rating

plt.pie(ratings,
        labels=ratings.index,
        colors=['magenta','cyan','green','red'], startangle=90,
        shadow= True ,autopct = '%1.1f%%', explode = [0,0,0,0.2,]

       )
plt.title('Movie ratings')
plt.legend(loc = 'lower left');
```



CORRELATION HEATMAP PLOT OF THE NETFLIX DATA

[78]:
```
#Plots numerical data as color encoded matrix

sns.heatmap(netflix.corr())
plt.title('Heatmap')
#plt.savefig('Heatmap.png', dpi= 80)
```

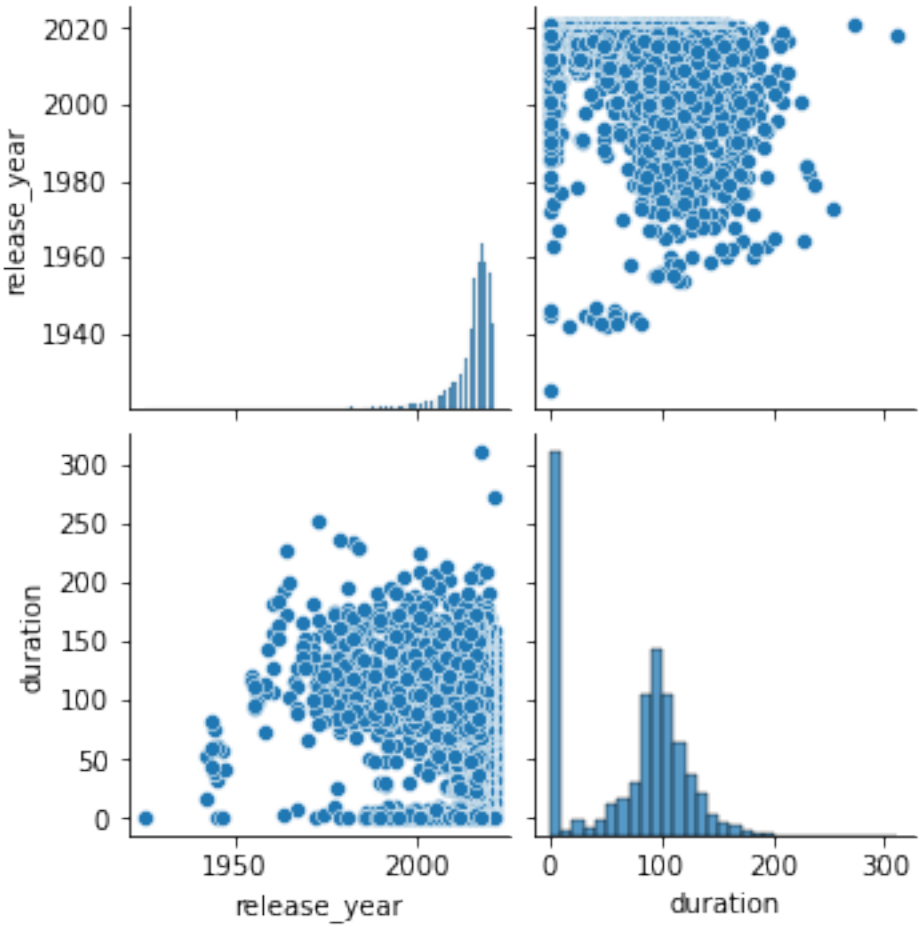
[78]:
```
Text(0.5, 1.0, 'Heatmap')
```



**** Pairplot plots pair wise relationship in a dataset****

[79]:
```
# Pairplot plots pair wise relationship in a dataset

sns.pairplot(netflix)
```

[79]: <seaborn.axisgrid.PairGrid at 0x215dc5e6370>



[80]:
```
netflix_tvshow = netflix[netflix['type'] == 'TV Show']
netflix_tvshow.head()
```

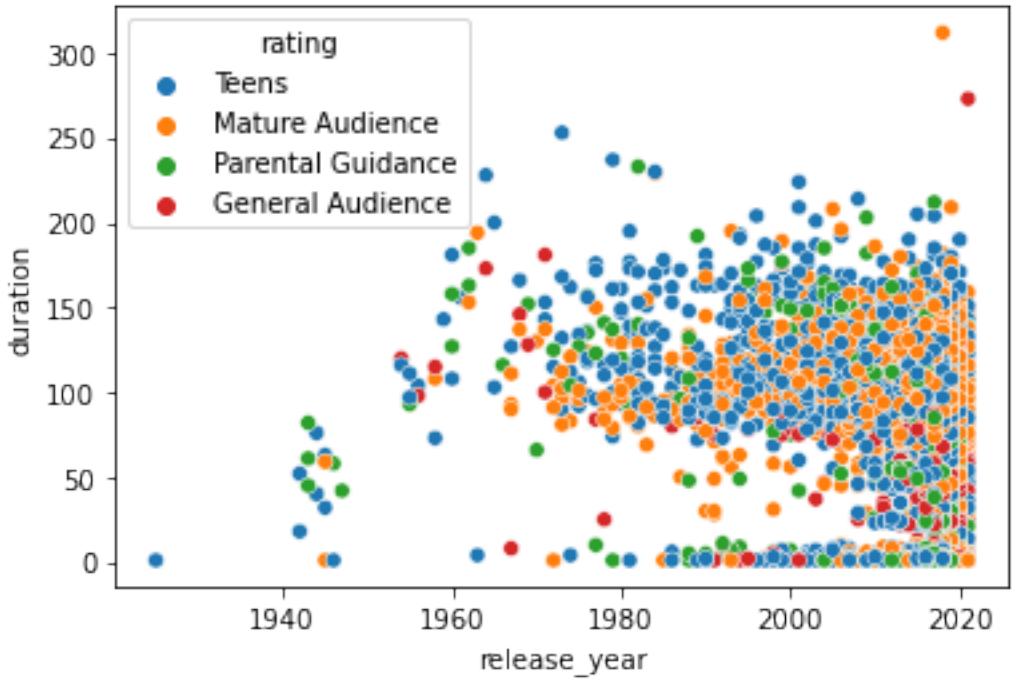
[80]:

	type	title	director	cast	country
1	TV Show	Blood & Water	None Specified	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…	South Africa
2	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi…	Not Specified
3	TV Show	Jailbirds New Orleans	None Specified	Not Listed	Not Specified
4	TV Show	Kota Factory	None Specified	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K…	India
5	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H…	Not Specified

```
          date_added  release_year           rating  duration  \
1  September 24, 2021          2021  Mature Audience         2
2  September 24, 2021          2021  Mature Audience         1
3  September 24, 2021          2021  Mature Audience         1
4  September 24, 2021          2021  Mature Audience         2
5  September 24, 2021          2021  Mature Audience         1

                                        listed_in  \
1    International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act…
3                             Docuseries, Reality TV
4  International TV Shows, Romantic TV Shows, TV …
5                 TV Dramas, TV Horror, TV Mysteries

                                         description
1  After crossing paths at a party, a Cape Town t…
2  To protect his family from a powerful drug lor…
3  Feuds, flirtations and toilet talk go down amo…
4  In a city of coaching centers known to train I…
5  The arrival of a charismatic young priest brin…
```

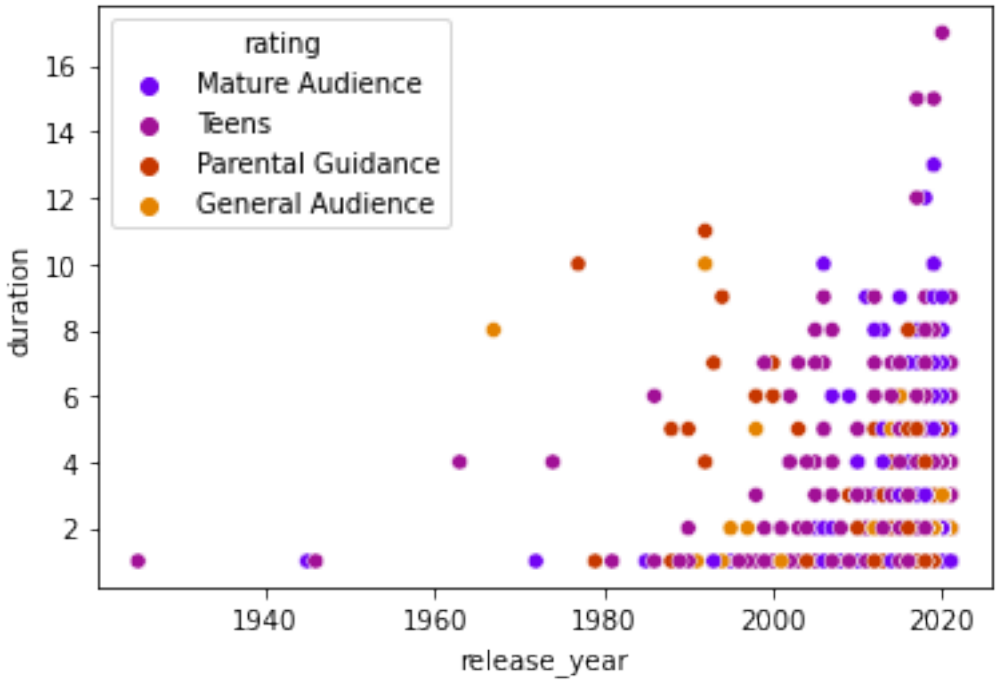
SCATTER PLOT OF RELEASE YEAR VERSUS DURATION OF MOVIE

[81]:
```
sns.scatterplot(x= 'release_year', y='duration', hue = 'rating', data =netflix);
```



***** DISTRIBUTION PLOT OF TV SHOW DURATION VERSUS RELEASE YEAR*****

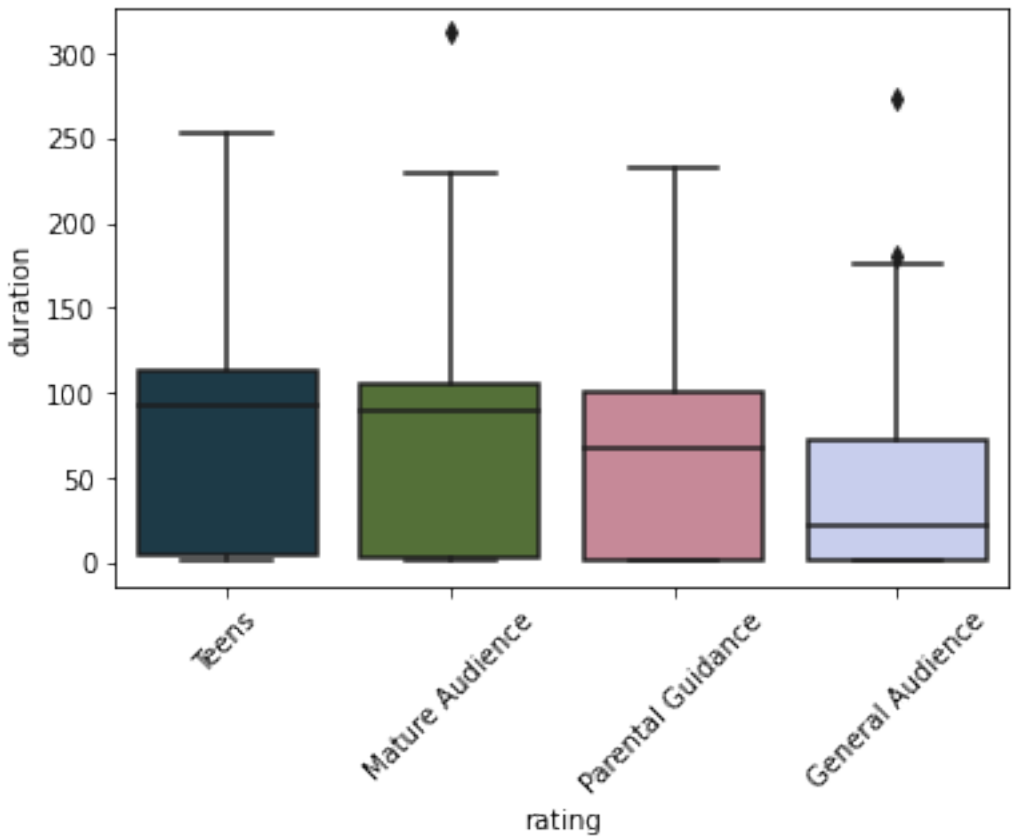
[82]:
```
#Distribution of TV show duration and release year
sns.scatterplot(x= 'release_year', y='duration', hue = 'rating', data␣
 ↪=netflix_tvshow, palette = 'gnuplot');
```



BOXPLOT OF TITLE RATING VERSUS DURATION

[83]:
```
sns.boxplot(x="rating", y="duration", data=netflix,palette='cubehelix');
plt.xticks(rotation = 45)
```

[83]:
```
(array([0, 1, 2, 3]),
 [Text(0, 0, 'Teens'),
  Text(1, 0, 'Mature Audience'),
  Text(2, 0, 'Parental Guidance'),
  Text(3, 0, 'General Audience')])
```



PLOT SHOWING POPULAR CAST

[84]:
```
# Splitting each cast to allow for accurate counting
names = netflix['cast'].str.split(',', expand=True).stack()
# Sorting the values and picking the top 10
popular = names.value_counts().sort_values(ascending=False).head(10)
popular
```

[84]:
```
Not Listed          825
 Anupam Kher         39
 Rupa Bhimani        31
 Takahiro Sakurai    30
 Julie Tejwani       28
 Om Puri             27
Shah Rukh Khan       26
 Rajesh Kava         26
 Paresh Rawal        25
 Yuki Kaji           25
dtype: int64
```

[85]:
```
popular.columns = ['Name']
popular = pd.DataFrame(popular)
popular.head(10)
```

[85]:

	0
Not Listed	825
Anupam Kher	39
Rupa Bhimani	31
Takahiro Sakurai	30
Julie Tejwani	28
Om Puri	27
Shah Rukh Khan	26
Rajesh Kava	26
Paresh Rawal	25
Yuki Kaji	25

BIVARIATE ANALYSIS

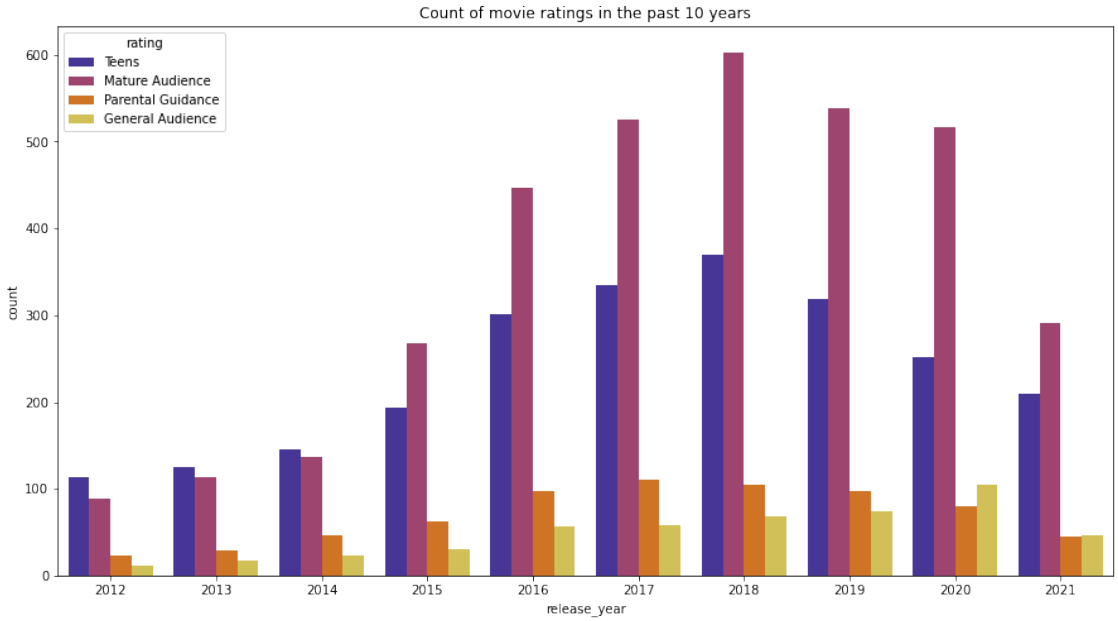
[88]:
```
# Selecting two columns to perform a bi-variate analysis
netflix_rating = netflix[['rating','release_year']]
# Selecting the last 10 years
last_10_years_rating = netflix_rating[netflix['release_year'] >= 2012]
last_10_years_rating
```

[88]:

	rating	release_year
0	Teens	2020
1	Mature Audience	2021
2	Mature Audience	2021
3	Mature Audience	2021
4	Mature Audience	2021
…	…	…
8798	Mature Audience	2014
8800	Parental Guidance	2012
8801	Mature Audience	2015
8803	Teens	2018
8806	Teens	2015

[7081 rows x 2 columns]

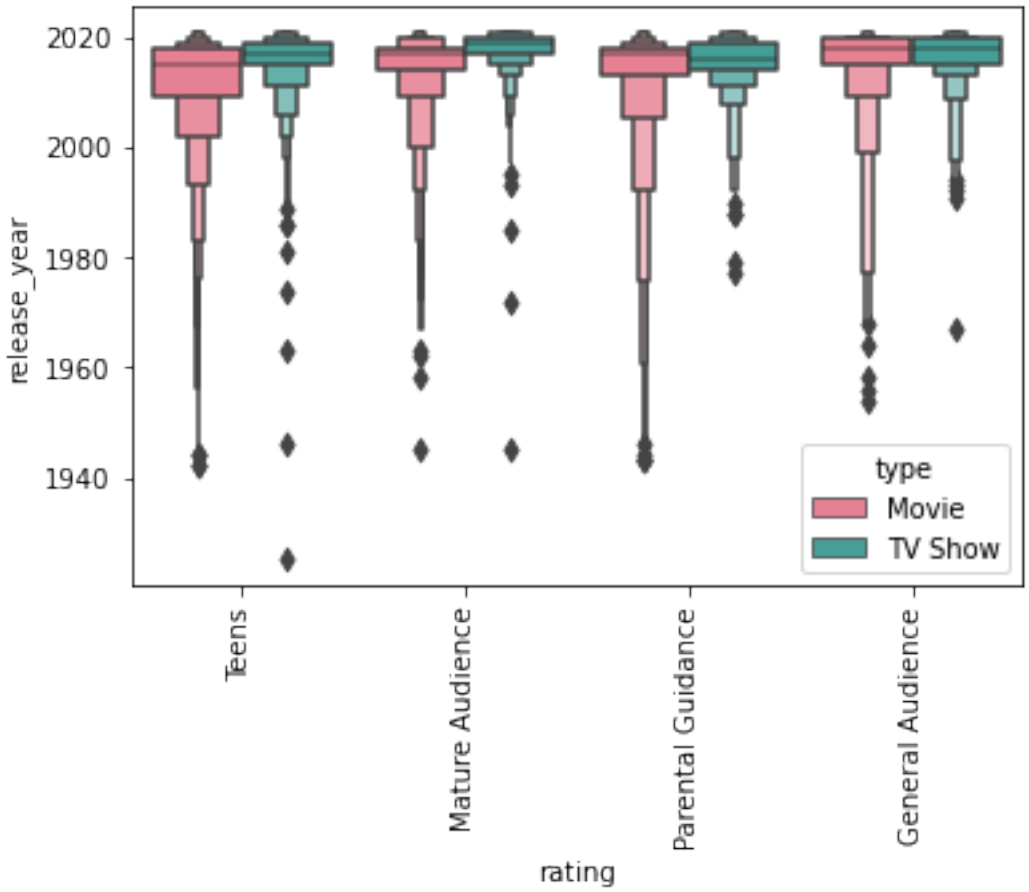
[89]:
```
# Bar plot
plt.figure(figsize=(15, 8))
sns.countplot(x='release_year', data=last_10_years_rating, hue='rating',␣
 ↪palette='CMRmap')
plt.title('Count of movie ratings in the past 10 years');
```



BIVARAITE ANALYSIS OF TITLE RATING VERSUS RELEASE YEAR

[90]:
```
sns.boxenplot(x='rating', y='release_year', hue='type', data = netflix ,␣
↪palette = 'husl')
plt.xticks(rotation = 90)
```

[90]:
```
(array([0, 1, 2, 3]),
 [Text(0, 0, 'Teens'),
  Text(1, 0, 'Mature Audience'),
  Text(2, 0, 'Parental Guidance'),
  Text(3, 0, 'General Audience')])
```



VISUALIZING NETFLIX TOP FIVE TITLE RATING AGAINST THE RELEASE YEAR

[101]:
```
netflix_rating = netflix[['rating', 'release_year']]
```

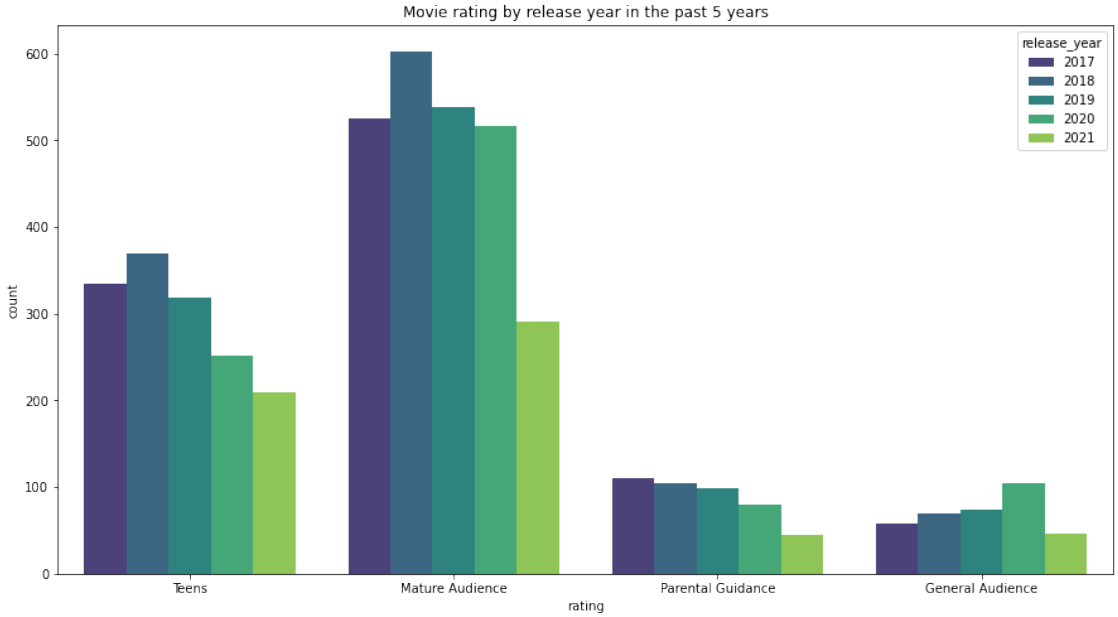
[102]:
```
# Selecting the last 5 years
last_5_years = netflix_rating[netflix['release_year'] >= 2017]
last_5_years
```

[102]:

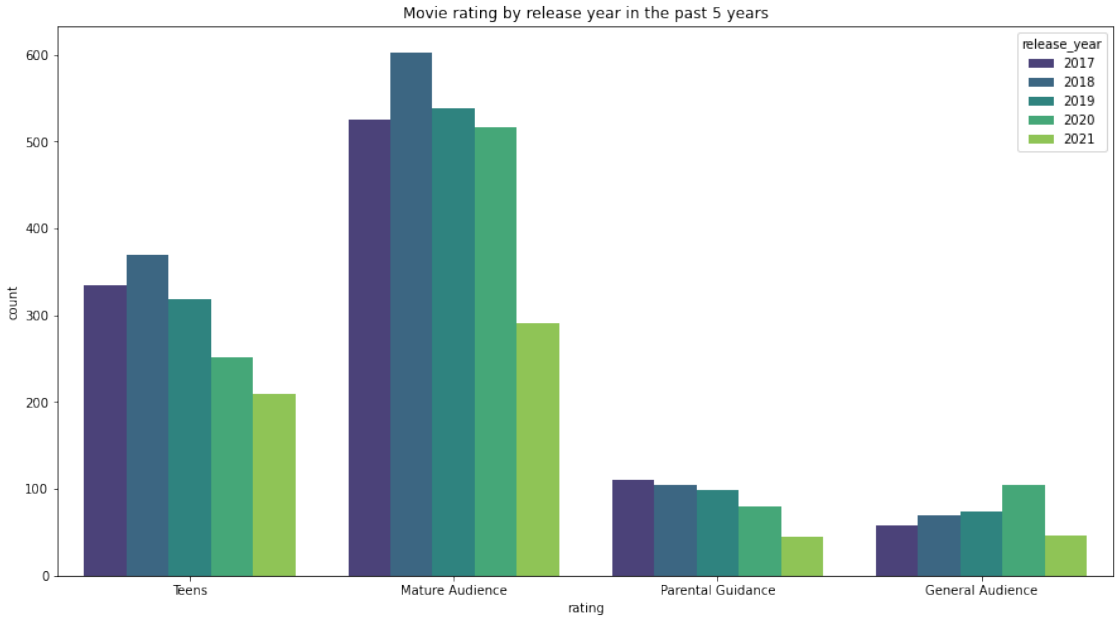
	rating	release_year
0	Teens	2020
1	Mature Audience	2021
2	Mature Audience	2021
3	Mature Audience	2021
4	Mature Audience	2021
…	…	…
8774	Teens	2018
8775	Parental Guidance	2018
8786	General Audience	2017

```
8787  Parental Guidance          2017
8803               Teens          2018

[4751 rows x 2 columns]
```

[103]:
```
plt.figure(figsize=(15, 8))
sns.countplot(x='rating', data=last_5_years, hue='release_year',␣
 ↪palette='viridis')
plt.title('Movie rating by release year in the past 5 years');
```



[104]:
```
plt.figure(figsize=(15, 8))
sns.countplot(x='rating', data=last_5_years, hue='release_year',␣
 ↪palette='viridis')
plt.title('Movie rating by release year in the past 5 years');
```



CONTRIBUTORS Margaret Oluwadare

Loveth Osuagwu

Oluchi Okoro (Oluchi Oluchi)

Monsurat Onabajo

Mariam Anishere

Olubusayo Solola

Olayemi Ibiloye

Maryann Amaefula

Olayemi Oloyede

Oluwadunsin Olajide

Type Markdown and LaTeX: 2

[ ]: