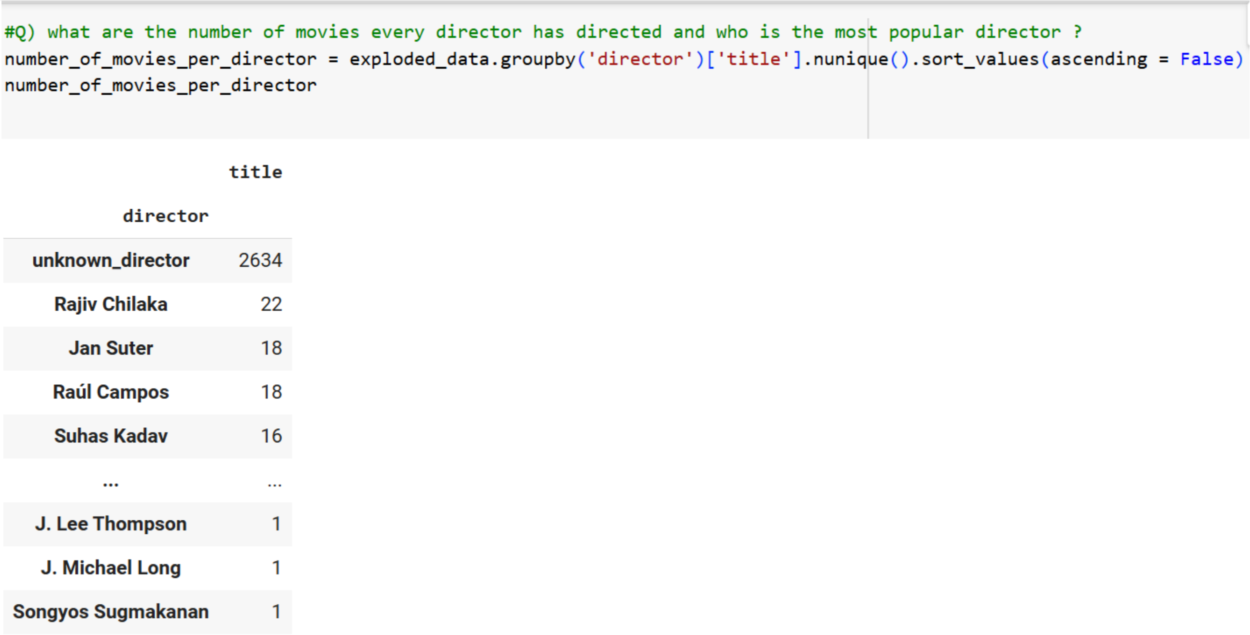
Q1) what are the number of movies every director has directed and who is the most popular director ?
Ans) `number_of_movies_per_director = exploded_data.groupby('director')['title'].nunique().sort_values(ascending = False)`

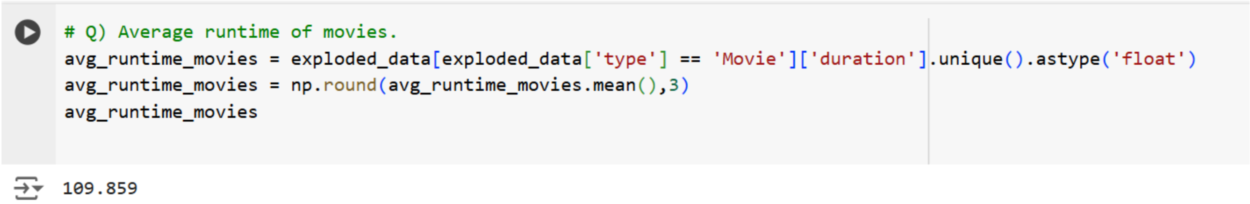


```
#Q) what are the number of movies every director has directed and who is the most popular director ?
number_of_movies_per_director = exploded_data.groupby('director')['title'].nunique().sort_values(ascending = False)
number_of_movies_per_director
```

	title
director	
unknown_director	2634
Rajiv Chilaka	22
Jan Suter	18
Raúl Campos	18
Suhas Kadav	16
...	...
J. Lee Thompson	1
J. Michael Long	1
Songyos Sugmakanan	1

Insights : As per query, there are directors who has directed movies to numbers as high as 22 to as low as 1. Though it should also be noted that due to data unavailability director names for lot of the movies are not recorded. Hence collectively all those values are shown as unknown_director which tops the list. But with the given amount of information, Rajiv Chilaka is the most popular director who has directed 22 movies up untill now.

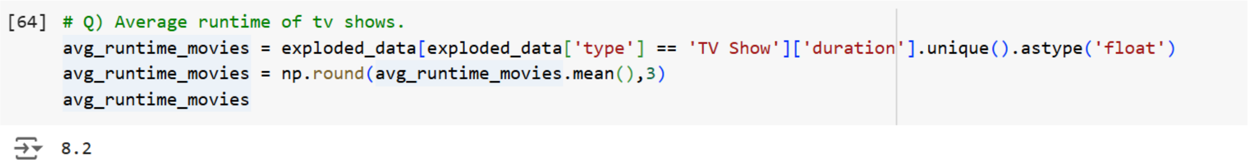
Q2) What is the average runtime of movies available in netflix.
Ans)



```
# Q) Average runtime of movies.
avg_runtime_movies = exploded_data[exploded_data['type'] == 'Movie']['duration'].unique().astype('float')
avg_runtime_movies = np.round(avg_runtime_movies.mean(),3)
avg_runtime_movies
```

```
109.859
```

Insights : The average run time of the movies available in the netflix is 109.859 minutes. This means that less than 2 hours is the average movie timings that is available. This gives the insight that the most likely the customers would like movie which are less than or equal to 2 hours. Anything more than the average value is not so welcomed by majority of the customers.
Recommendations : Netflix should source more movies which are less than 2 hours in run time.
Q3) What is average runtime for TV shows ?
Ans)



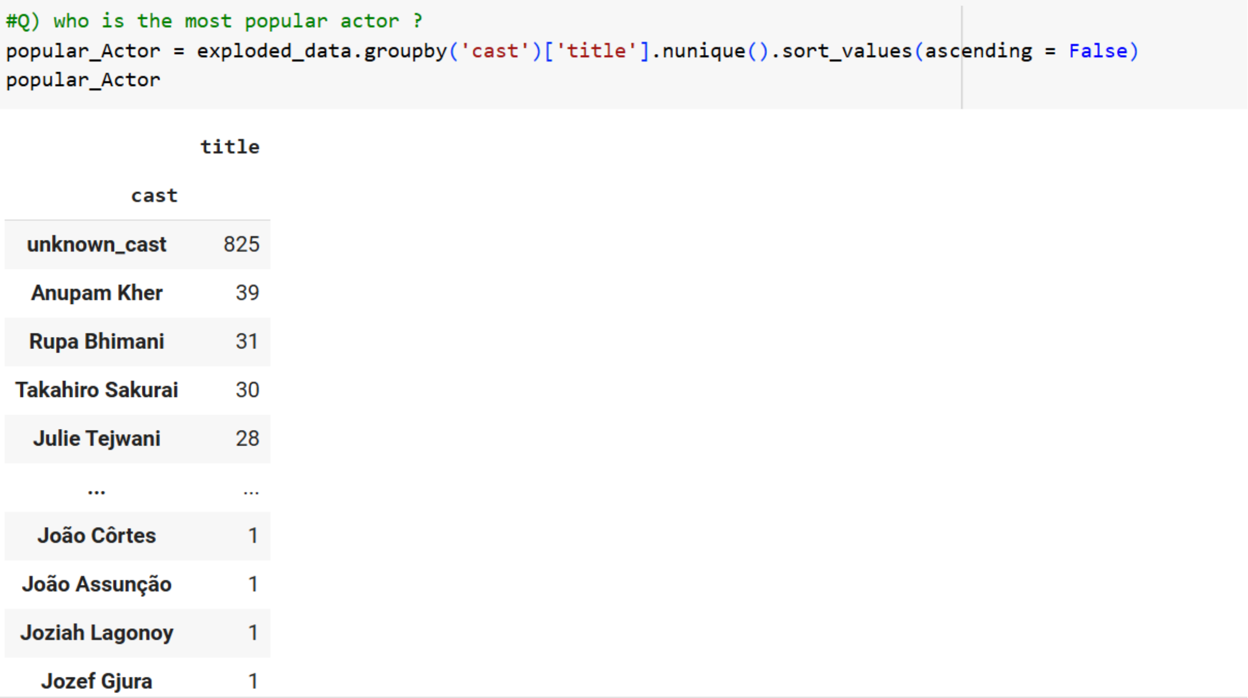
```
[64] # Q) Average runtime of tv shows.
     avg_runtime_movies = exploded_data[exploded_data['type'] == 'TV Show']['duration'].unique().astype('float')
     avg_runtime_movies = np.round(avg_runtime_movies.mean(),3)
     avg_runtime_movies
```

```
8.2
```

Insights : The Average number of seasons that are present in netflix for the tv shows are 8. This means that generally people are preferring the tv shows which are long and engaging. 8 seasons would definitely be giving enough time for the viewers to have an emotional connect to the characters present in the TV shows. Hence long format tv shows are more likely to be watched by the viewers.

Recommendations : Netflix can source more Tv shows which are in general longer in nature. A sweet spot should be anything less than 8. But make sure that it is not more than 8 seasons as that might be felt as dragging for the viewers and thus reduce the viewership.

Q4) Who is the most popular Actor ?
Ans)



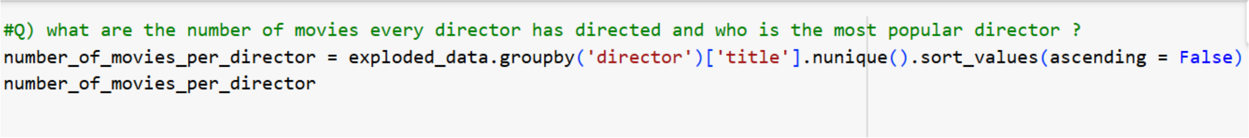
```
#Q) who is the most popular actor ?
popular_Actor = exploded_data.groupby('cast')['title'].nunique().sort_values(ascending = False)
popular_Actor
```

	title
cast	
unknown_cast	825
Anupam Kher	39
Rupa Bhimani	31
Takahiro Sakurai	30
Julie Tejwani	28
...	...
João Côrtes	1
João Assunção	1
Joziah Lagonoy	1
Jozef Gjura	1

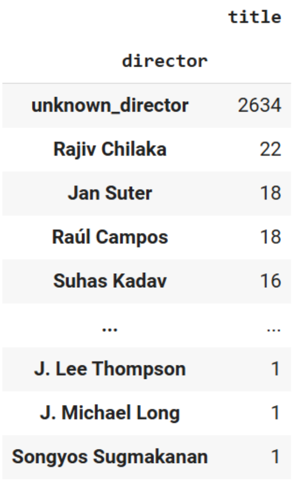
Insights : The most popular actor based on the number of movies they have acted is Anupam Kher. He has acted in 39 movies. There are 825 unknown cast which tops the list. But these are the cast which are not available. So we can ignore them and consider Anupam Kher as the most popular actor .

Recommendations : Movies of Anupam Kher are more likely to be seen by the viewers. Hence weightage can be given when sourcing movies which are acted by Anupam Kher.

Q5) Who is the most popular Director ?
Ans)



```
#Q) what are the number of movies every director has directed and who is the most popular director ?
number_of_movies_per_director = exploded_data.groupby('director')['title'].nunique().sort_values(ascending = False)
number_of_movies_per_director
```

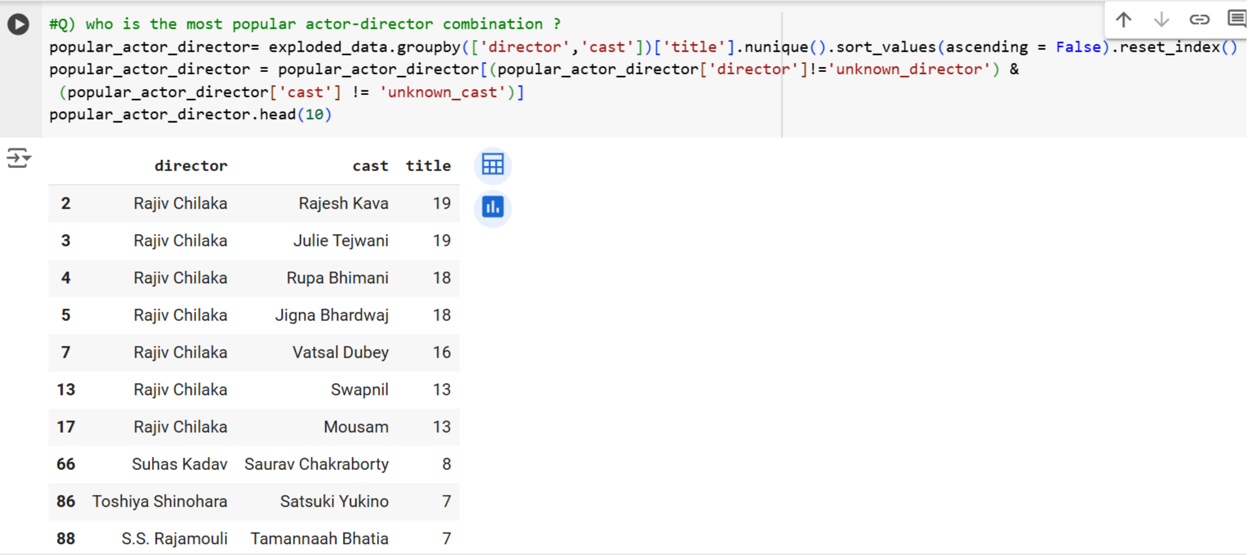


	title
director	
unknown_director	2634
Rajiv Chilaka	22
Jan Suter	18
Raúl Campos	18
Suhas Kadav	16
...	...
J. Lee Thompson	1
J. Michael Long	1
Songyos Sugmakanan	1

Insights : The most popular director is Rajiv Chilaka who has director 22 movies. There are other data whose director is unknown. Those are counted as Unknown director which sums up to 2634. But among the listed known items, Rajiv Chilaka is the most popular director.

Recommendations : Movies of Rajiv are more likely to be seen by the viewers. Hence weightage can be given when sourcing movies which are directed by Rajiv Chilaka

Q6) Which is the most popular actor – director combination?
Ans)



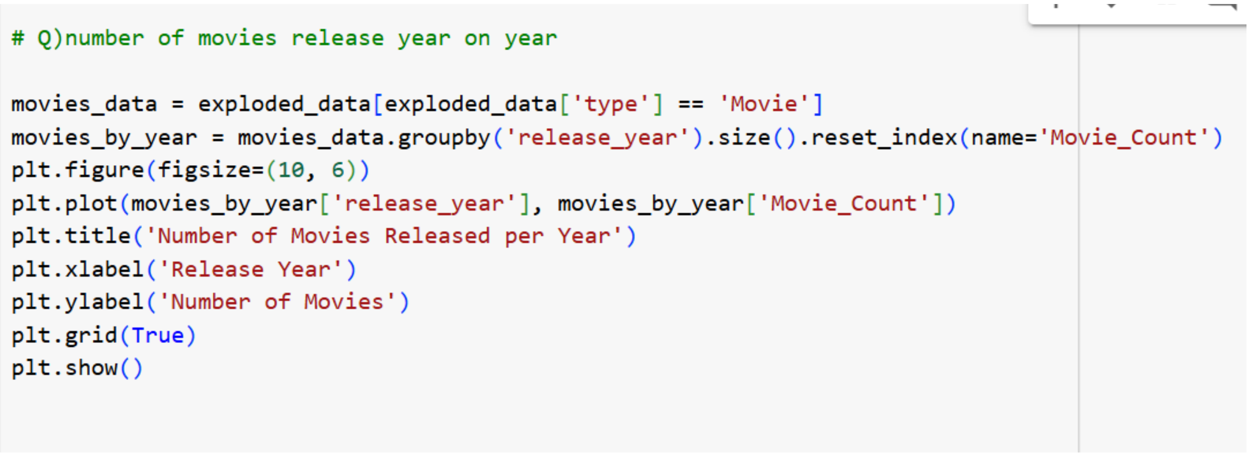
```
#Q) who is the most popular actor-director combination ?
popular_actor_director= exploded_data.groupby(['director','cast'])['title'].nunique().sort_values(ascending = False).reset_index()
popular_actor_director = popular_actor_director[(popular_actor_director['director']!='unknown_director') &
 (popular_actor_director['cast'] != 'unknown_cast')]
popular_actor_director.head(10)
```

	director	cast	title
2	Rajiv Chilaka	Rajesh Kava	19
3	Rajiv Chilaka	Julie Tejwani	19
4	Rajiv Chilaka	Rupa Bhimani	18
5	Rajiv Chilaka	Jigna Bhardwaj	18
7	Rajiv Chilaka	Vatsal Dubey	16
13	Rajiv Chilaka	Swapnil	13
17	Rajiv Chilaka	Mousam	13
66	Suhas Kadav	Saurav Chakraborty	8
86	Toshiya Shinohara	Satsuki Yukino	7
88	S.S. Rajamouli	Tamannaah Bhatia	7

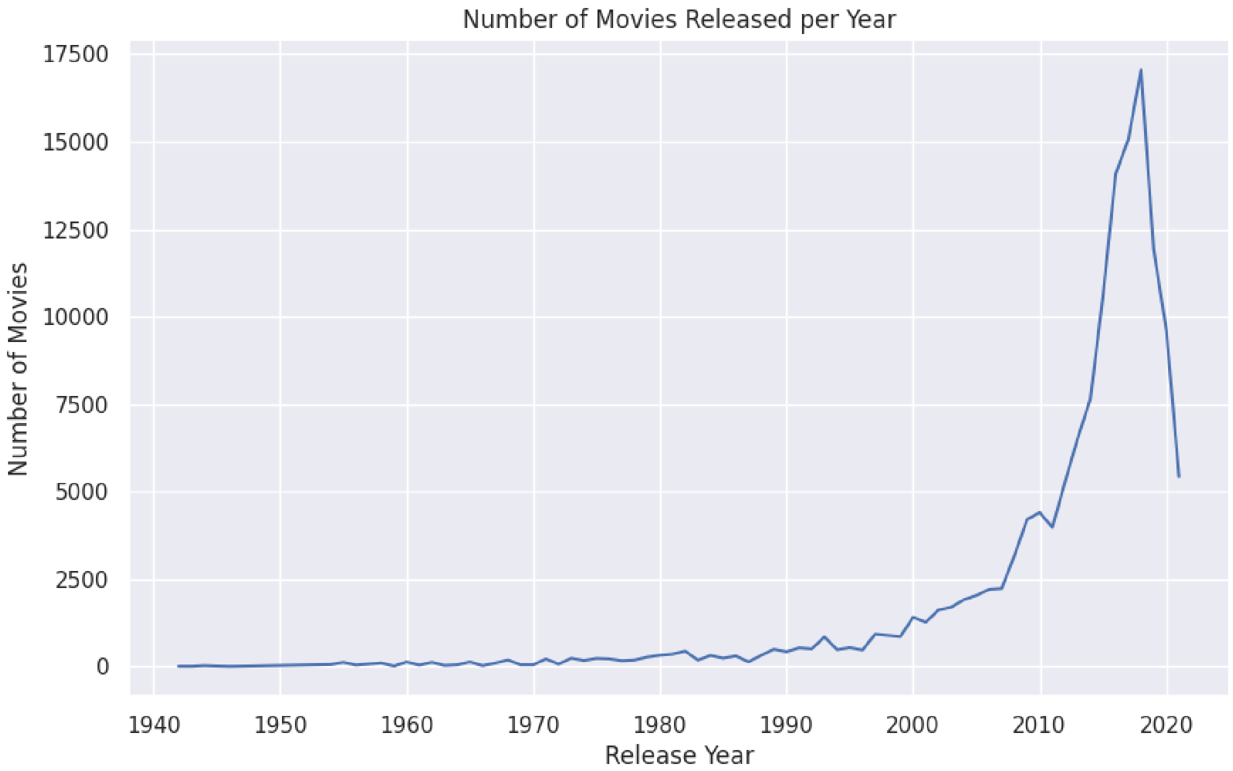
Insights : the most popular actor – director combination is Rajiv Chilaka and Rajesh Kava. Along with them there Rajiv Chilaka and Julie Tejwani.There are 19 movies which brings in these two actors along with same director combination. There are other actor who have worked with Rajiv like Rupa and Jigna who have a combination count of 18 which is as much equal as Rajesh and Julie.

Recommendations : The movies that have the actor director combination of Rajiv – Rajesh and Rajiv Julie can be given preference when sourcing content. Along with that we can also consider the combination Rajiv – Rupa and Rajiv – Jigna. These combinations are also as popular as the top most one.

Q7) What is the trend of movies releases over the years
Ans)



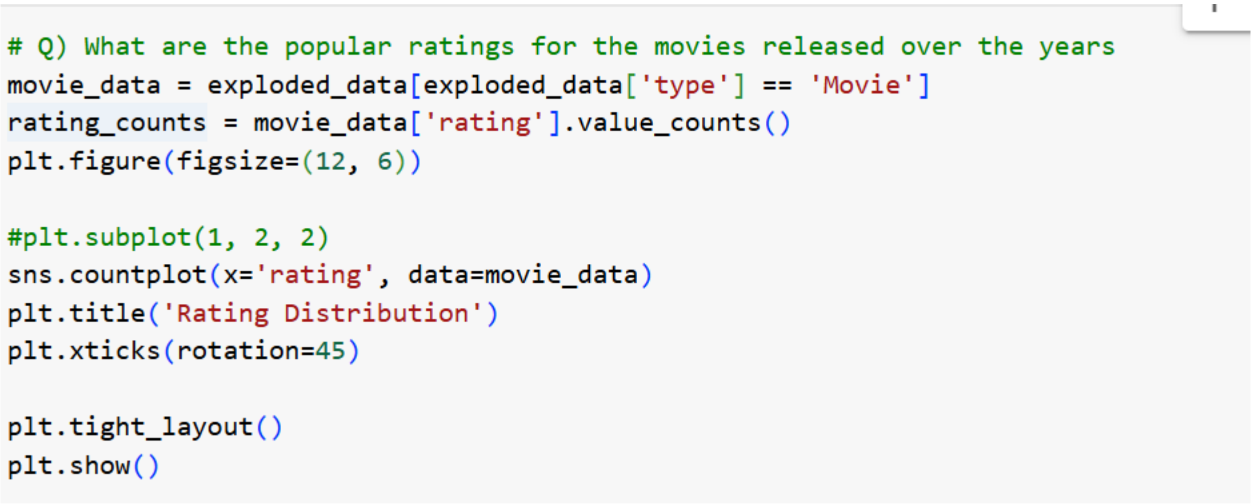
```
# Q)number of movies release year on year

movies_data = exploded_data[exploded_data['type'] == 'Movie']
movies_by_year = movies_data.groupby('release_year').size().reset_index(name='Movie_Count')
plt.figure(figsize=(10, 6))
plt.plot(movies_by_year['release_year'], movies_by_year['Movie_Count'])
plt.title('Number of Movies Released per Year')
plt.xlabel('Release Year')
plt.ylabel('Number of Movies')
plt.grid(True)
plt.show()
```



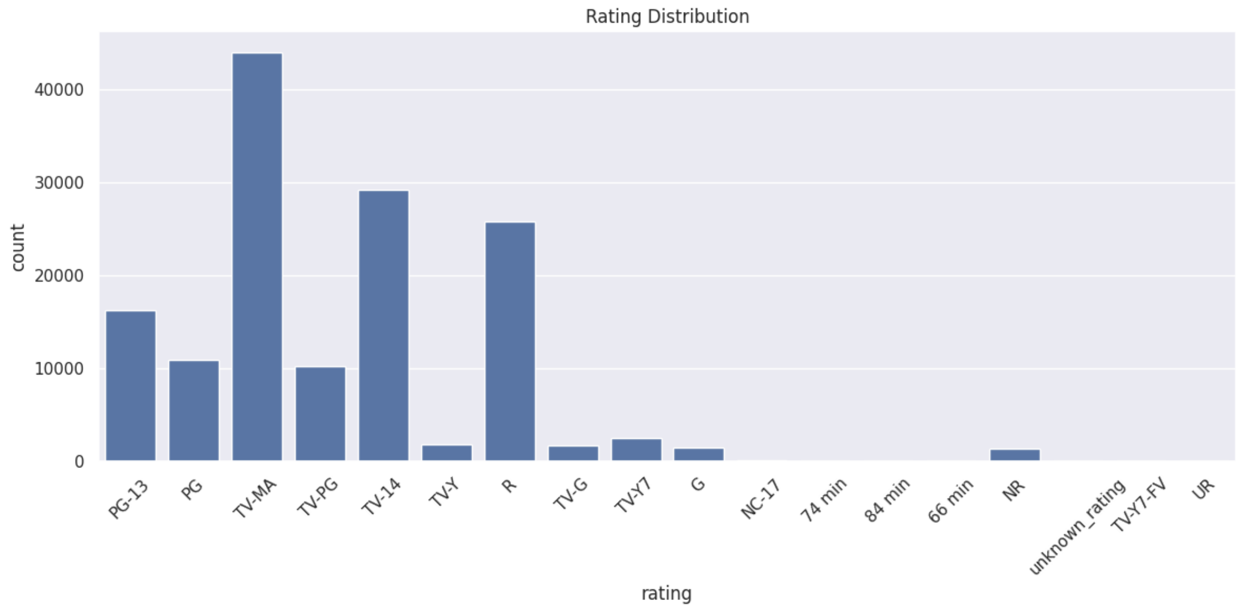
Insights : We can see that the number of movies released over the years has been increasing steadily but slow from 1940s to late 1990. After 2000, the number of movies getting released has been increased and there is a surge in movie releases during 2016 – 2018 period. After that there is a decrease in number of movies released. This can also be claimed for the pandemic that resulted in the complete shut down of outdoor activities.

Recommendations : The movie watching habit of people have been increasing steadily throughout these years. Though there is a dip in the number of movies due to pandemic, people are always ready to watch the movies out there. Also during the pandemic it was the OTT platforms that helped people to be engaged. So Netflix can definitely source more contents which would be increasing their viewer base as well as average view time.

Q8) Which are the popular ratings of the movies. ?
Ans)



```
# Q) What are the popular ratings for the movies released over the years
movie_data = exploded_data[exploded_data['type'] == 'Movie']
rating_counts = movie_data['rating'].value_counts()
plt.figure(figsize=(12, 6))

#plt.subplot(1, 2, 2)
sns.countplot(x='rating', data=movie_data)
plt.title('Rating Distribution')
plt.xticks(rotation=45)

plt.tight_layout()
plt.show()
```

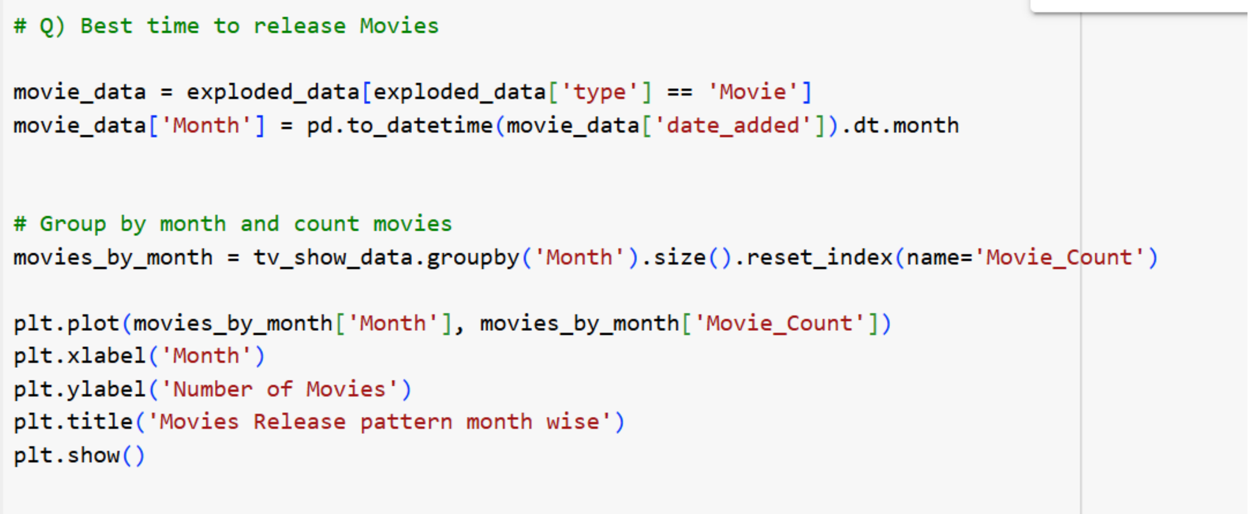


Insights: The popular rating of the movies as per graph says that TV_MA which have over 40000 count. The Second most popular rating for the movie that customers watch is TV-14 category whose counts are around 30000. The rating R is third position with similar count as that of TV 14 Category. There are other category where there are very veryless movie count too like NC-17, UR etc. Couple of the movies do not have rating info hence they are grouped under unknown_rating tag. This gives the insight that mature audience form majority of the customer base. The second most viewed category being TV-14 says that there are viewers who are as low as 14 years old.

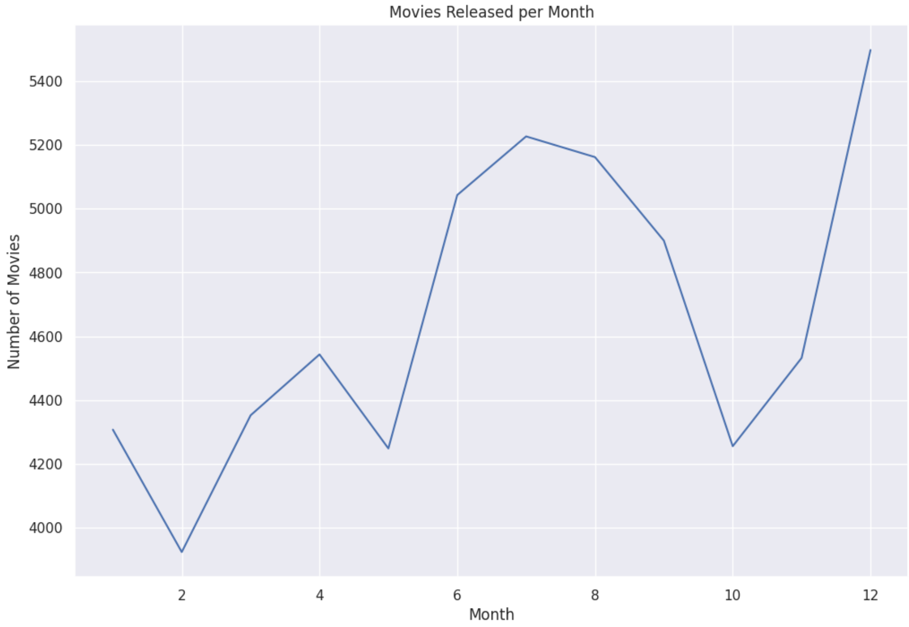
Recommendations : Netflix can give preference over TV_MA rated movies which are the most preferred one.Along with that the other rating that could improve the viewership would be TV-

14 and R. More content could be garnered for TV-14 though they are second in position. But good quality content if served to these cateogry, this would help in creating long term relationship with customers. And Since 14 years and above the major youth age group they would help spreading the appreciation of the platform , especially in social media and via word of mouth. Along with that we can also cater to the Mature audience who are already loyal to the platform.

Q9) What is the best time to release a movie.
Ans)



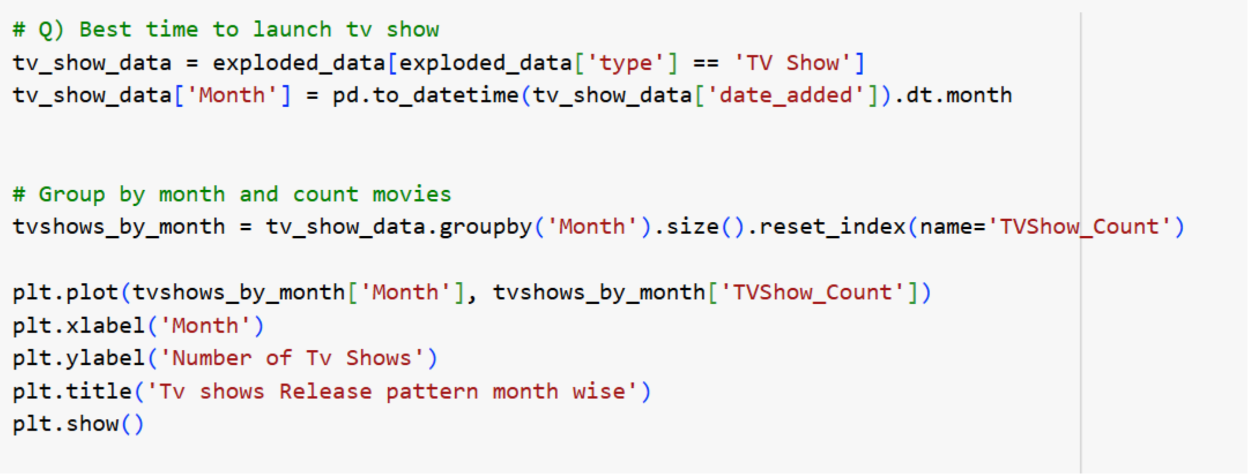
```
# Q) Best time to release Movies

movie_data = exploded_data[exploded_data['type'] == 'Movie']
movie_data['Month'] = pd.to_datetime(movie_data['date_added']).dt.month


# Group by month and count movies
movies_by_month = tv_show_data.groupby('Month').size().reset_index(name='Movie_Count')

plt.plot(movies_by_month['Month'], movies_by_month['Movie_Count'])
plt.xlabel('Month')
plt.ylabel('Number of Movies')
plt.title('Movies Release pattern month wise')
plt.show()
```



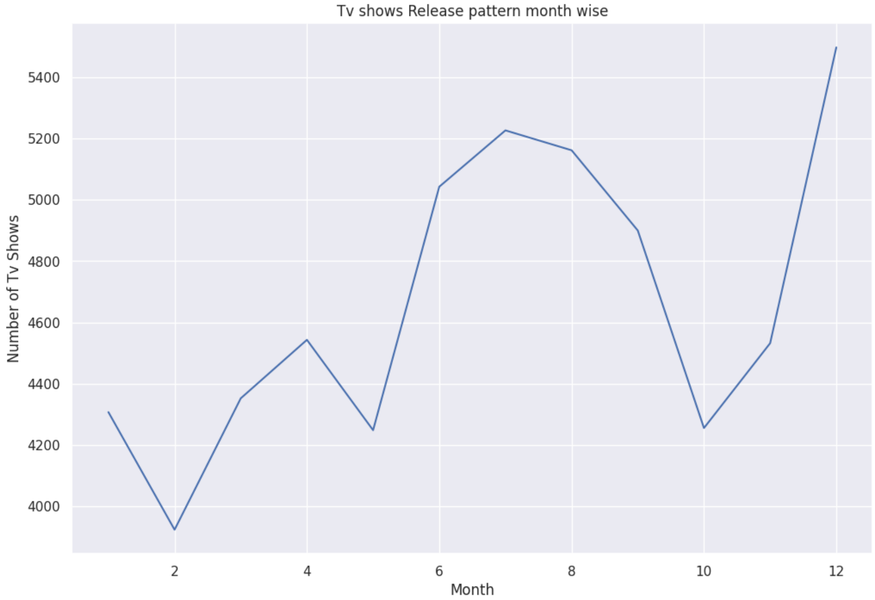
Insights : As per the data available in the platform, the best time to release the movie would be 12th month i.e. December month of the year. This is the time for vacation and celebration and people would want to celebrate being together by watching a movie. The second best time to release the movie is 7th Month which is July. These are the two months where maximum movies were released across all these years.

Recommendations : More movies could be sourced for the December month so that users would be able to watch them during the holiday season. The mid year updation of the movie list is also another good time as to improve the customer base and viewership time.

Q10) what is the best time to release a TV show ?
Ans)



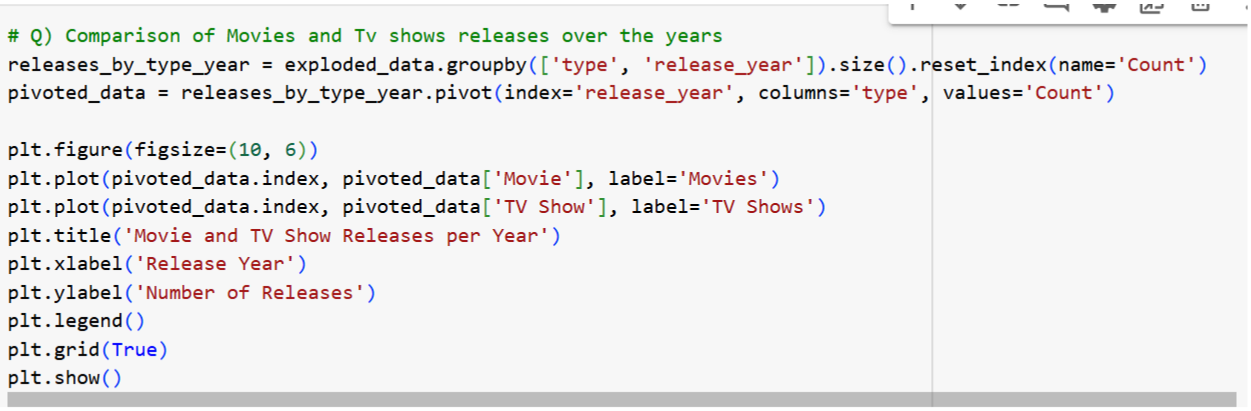
```
# Q) Best time to launch tv show
tv_show_data = exploded_data[exploded_data['type'] == 'TV Show']
tv_show_data['Month'] = pd.to_datetime(tv_show_data['date_added']).dt.month


# Group by month and count movies
tvshows_by_month = tv_show_data.groupby('Month').size().reset_index(name='TVShow_Count')

plt.plot(tvshows_by_month['Month'], tvshows_by_month['TVShow_Count'])
plt.xlabel('Month')
plt.ylabel('Number of Tv Shows')
plt.title('Tv shows Release pattern month wise')
plt.show()
```



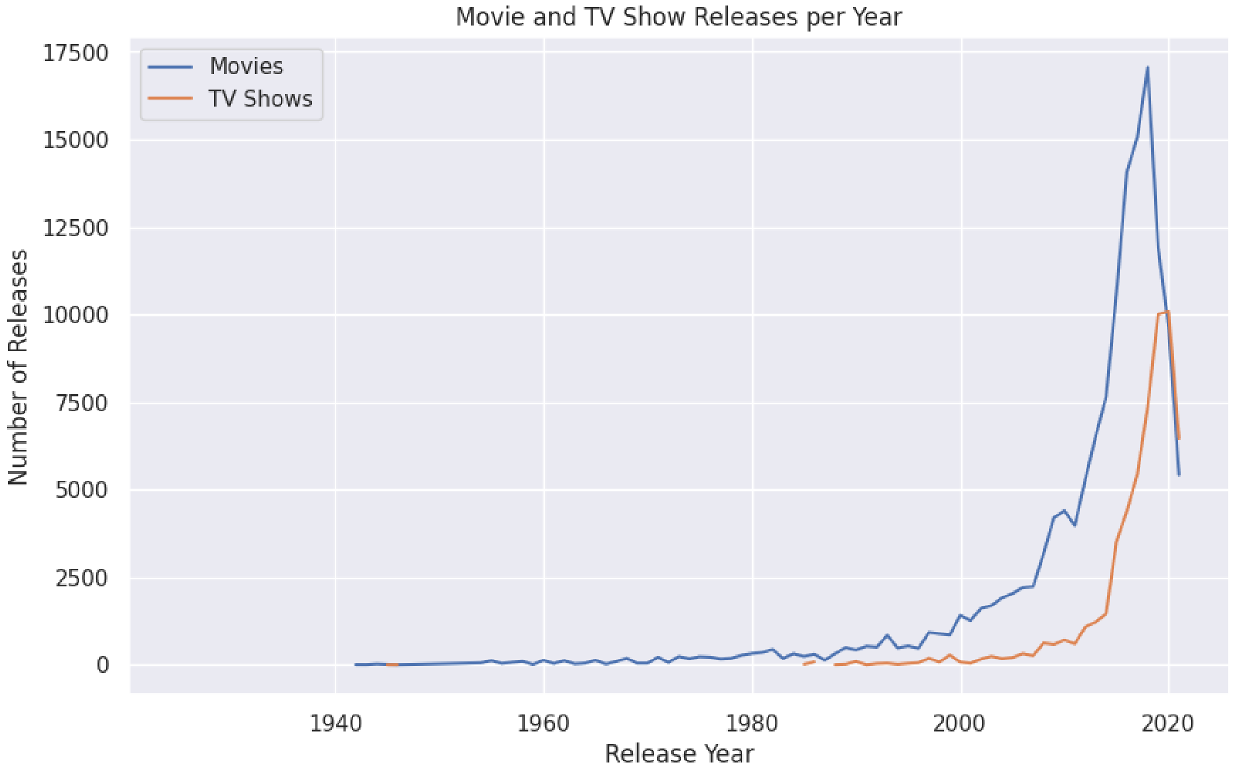
Insights : The best time to launch TV show is December and July just like the movies. Since its the holiday season people have the time to get engaged with the online contents available.
Recommendations : More TV shows could be acquired prior to the peak holiday season as to get more viewership

Q11) Compare the releases of Movies and TV shows over the period of time ?

Ans)



```
# Q) Comparison of Movies and Tv shows releases over the years
releases_by_type_year = exploded_data.groupby(['type', 'release_year']).size().reset_index(name='Count')
pivoted_data = releases_by_type_year.pivot(index='release_year', columns='type', values='Count')

plt.figure(figsize=(10, 6))
plt.plot(pivoted_data.index, pivoted_data['Movie'], label='Movies')
plt.plot(pivoted_data.index, pivoted_data['TV Show'], label='TV Shows')
plt.title('Movie and TV Show Releases per Year')
plt.xlabel('Release Year')
plt.ylabel('Number of Releases')
plt.legend()
plt.grid(True)
plt.show()
```

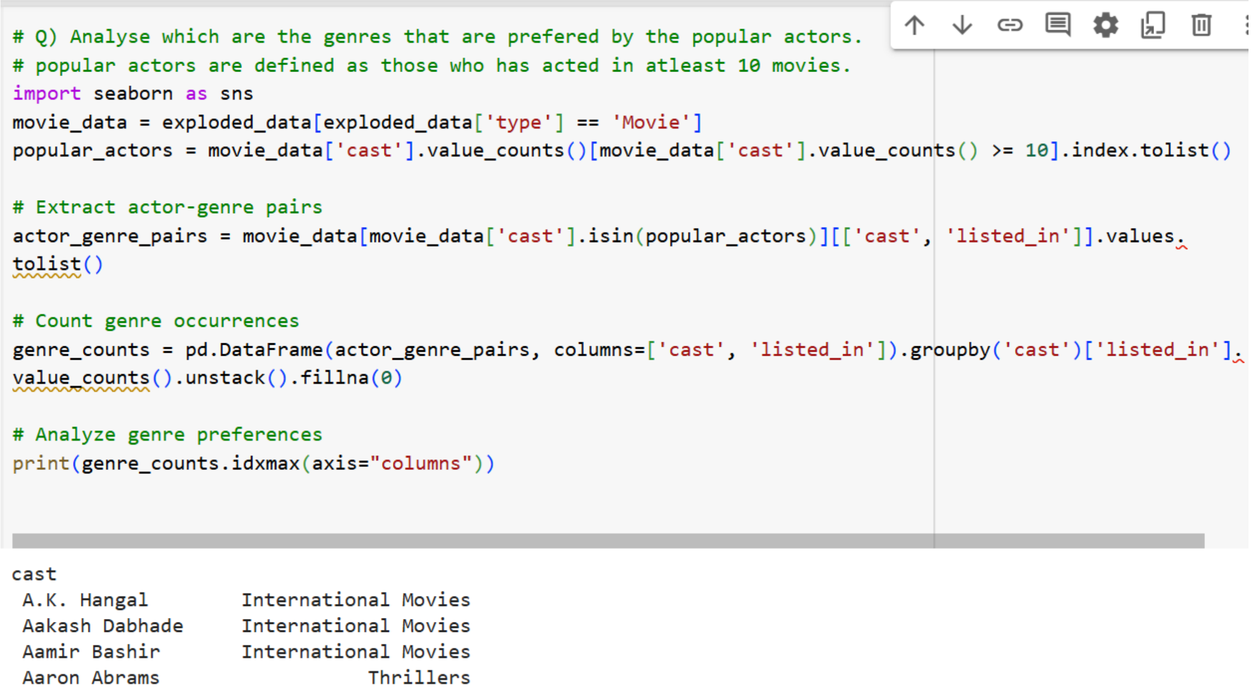


Insights: Over the period of time, there is increasing the number of tv shows and movies that are released. The movies count increased starting from early 2000 while the tv shows count increased in the late 2000s. During the period of 2016-2018 there is a stark increase in count of both movies and tv shows. In these, the movies count are more when compared to that of Tv shows. But yes this is obvious as the tv timings are always crucial for the releases and also the duration unlike movies. During the late 2019 there is decrease in releases of both movies and tv shows especially because of pandemic that was present.

But it is to be noted that the engagement in the platform has always been increasing especially since after onset of pandemic

Recommendations : Netflix when source should give importance to both tv shows as well as movies as their release counts has been increased in the recent years though they were a hit on the same due to the pandemic situation.

Q12) Analyse which are the genres preffered byt eh popular actors ?
Ans)



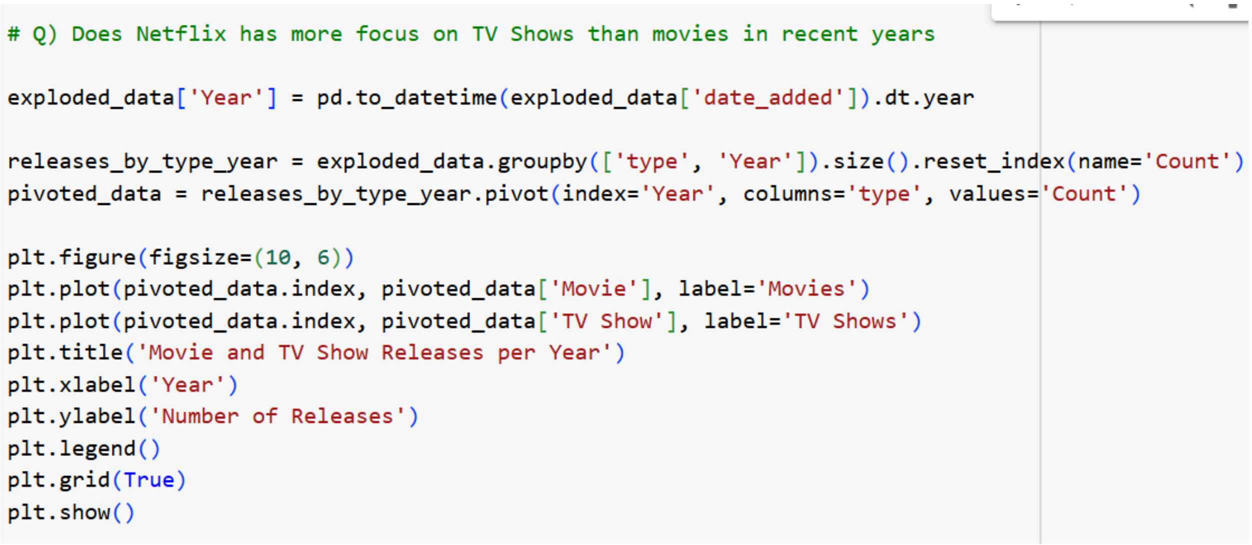
```
# Q) Analyse which are the genres that are prefered by the popular actors.
# popular actors are defined as those who has acted in atleast 10 movies.
import seaborn as sns
movie_data = exploded_data[exploded_data['type'] == 'Movie']
popular_actors = movie_data['cast'].value_counts()[movie_data['cast'].value_counts() >= 10].index.tolist()

# Extract actor-genre pairs
actor_genre_pairs = movie_data[movie_data['cast'].isin(popular_actors)][['cast', 'listed_in']].values.
tolist()

# Count genre occurrences
genre_counts = pd.DataFrame(actor_genre_pairs, columns=['cast', 'listed_in']).groupby('cast')['listed_in'].
value_counts().unstack().fillna(0)

# Analyze genre preferences
print(genre_counts.idxmax(axis="columns"))
```

```
cast
 A.K. Hangal          International Movies
 Aakash Dabhade       International Movies
 Aamir Bashir         International Movies
 Aaron Abrams                    Thrillers
```

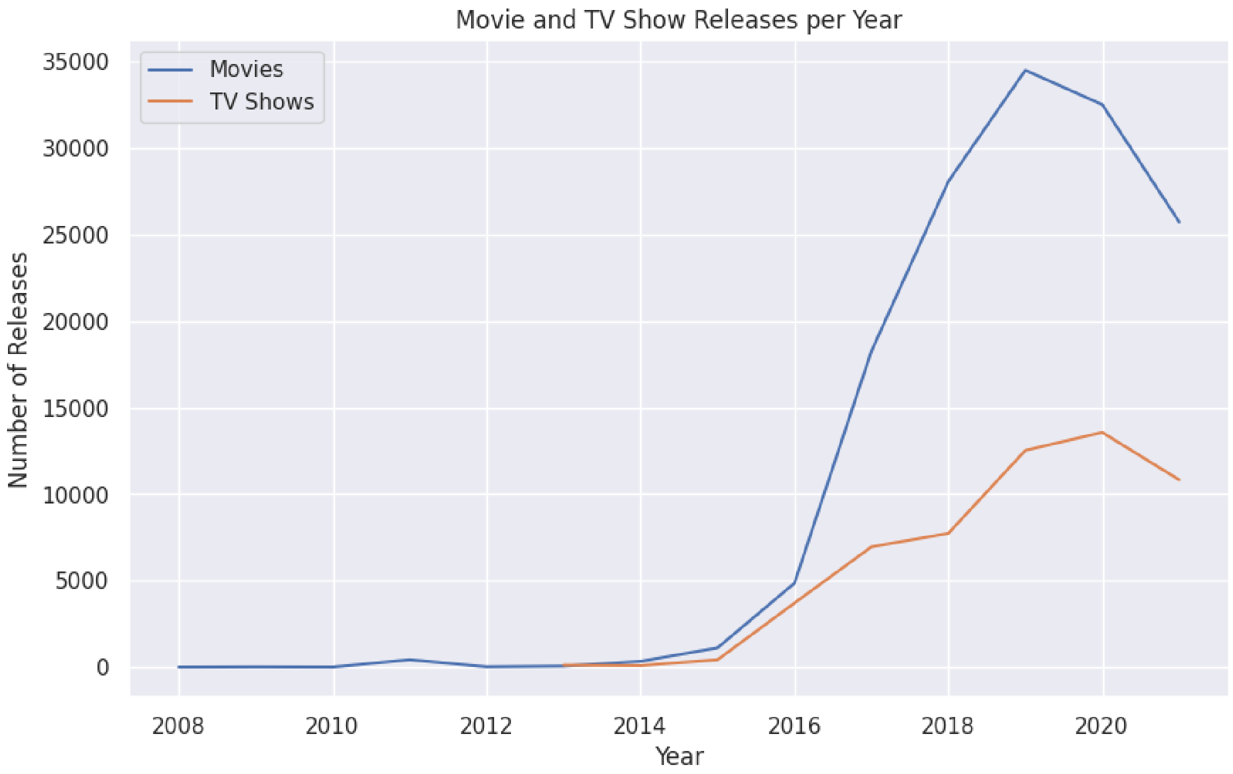
Insights : It can be seen that most of the popular actors (which are defined as actors who has a movie count of more than 10) has been interested in International Movies. Along with them there are other categories like thrillers, Dramas etc.

Recommendations : More movies can be added to the list which belongs to the Categories of International Movies, Thrillers, Dramas etc. It could also be noted that the Popular actors are also regional in nature, so movies of the popular actors in regional languages too could be sourced that would be improve the engagement in the platform .

Q13) Does netflix has more focus on TV shows than movies in recent years ?



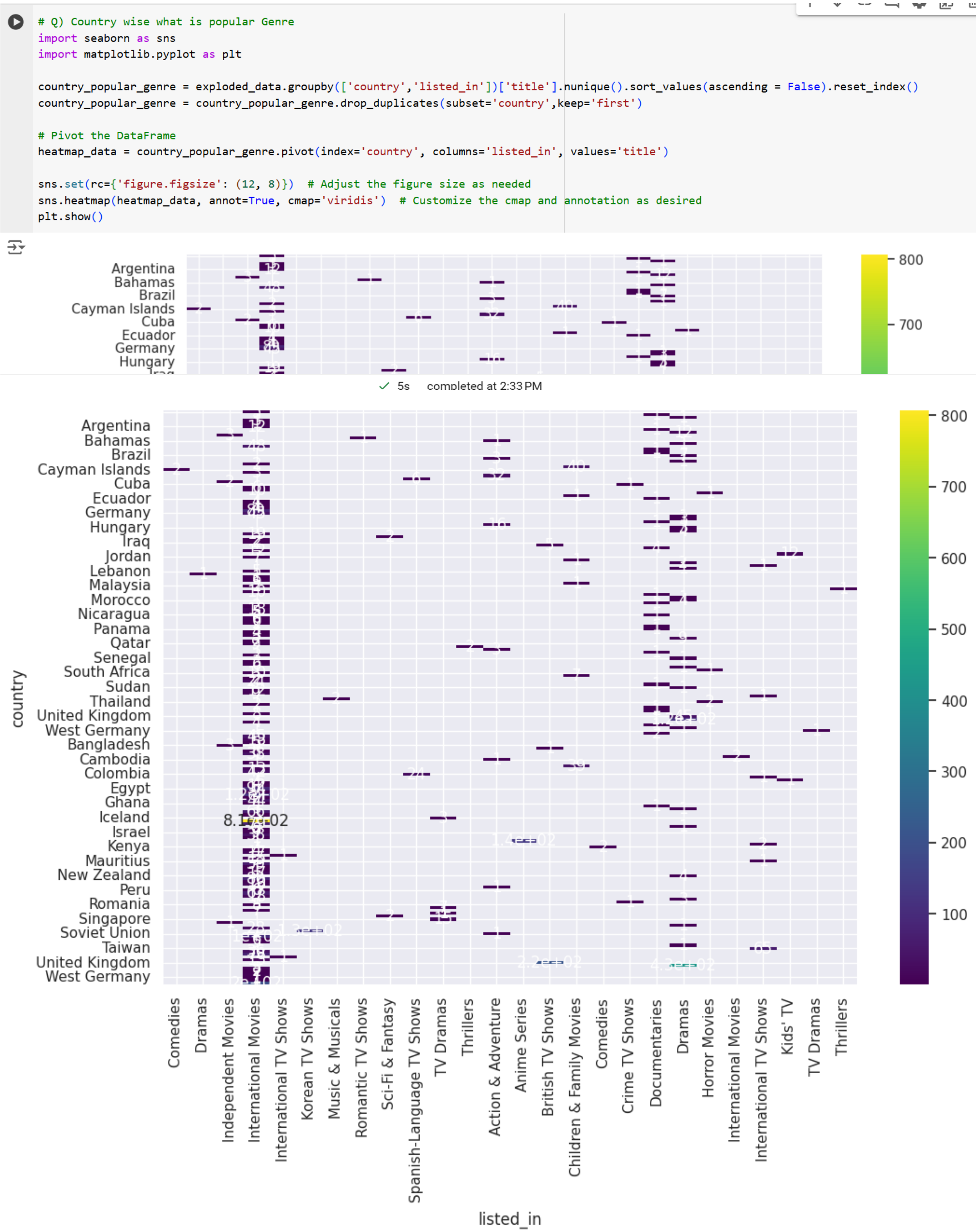
```python
# Q) Does Netflix has more focus on TV Shows than movies in recent years

exploded_data['Year'] = pd.to_datetime(exploded_data['date_added']).dt.year

releases_by_type_year = exploded_data.groupby(['type', 'Year']).size().reset_index(name='Count')
pivoted_data = releases_by_type_year.pivot(index='Year', columns='type', values='Count')

plt.figure(figsize=(10, 6))
plt.plot(pivoted_data.index, pivoted_data['Movie'], label='Movies')
plt.plot(pivoted_data.index, pivoted_data['TV Show'], label='TV Shows')
plt.title('Movie and TV Show Releases per Year')
plt.xlabel('Year')
plt.ylabel('Number of Releases')
plt.legend()
plt.grid(True)
plt.show()
```



Insights: No. Netflix has more focus on movies than the tv shows.It should also be noted there is steady increase in both movies and tv shows that are added. But More focus is always on Movies. Major reason being the long duration of the TV shows. The cool off period to be completed for TV shows are more when compared to that of Movies.

Q14) Visualise country wise popular genre ?

Ans)

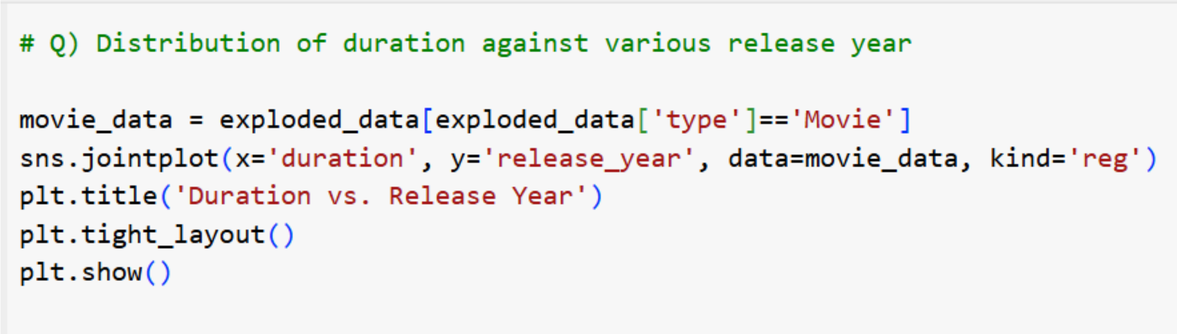


```
# Q) Country wise what is popular Genre
import seaborn as sns
import matplotlib.pyplot as plt

country_popular_genre = exploded_data.groupby(['country','listed_in'])['title'].nunique().sort_values(ascending = False).reset_index()
country_popular_genre = country_popular_genre.drop_duplicates(subset='country',keep='first')

# Pivot the DataFrame
heatmap_data = country_popular_genre.pivot(index='country', columns='listed_in', values='title')

sns.set(rc={'figure.figsize': (12, 8)})  # Adjust the figure size as needed
sns.heatmap(heatmap_data, annot=True, cmap='viridis')  # Customize the cmap and annotation as desired
plt.show()
```

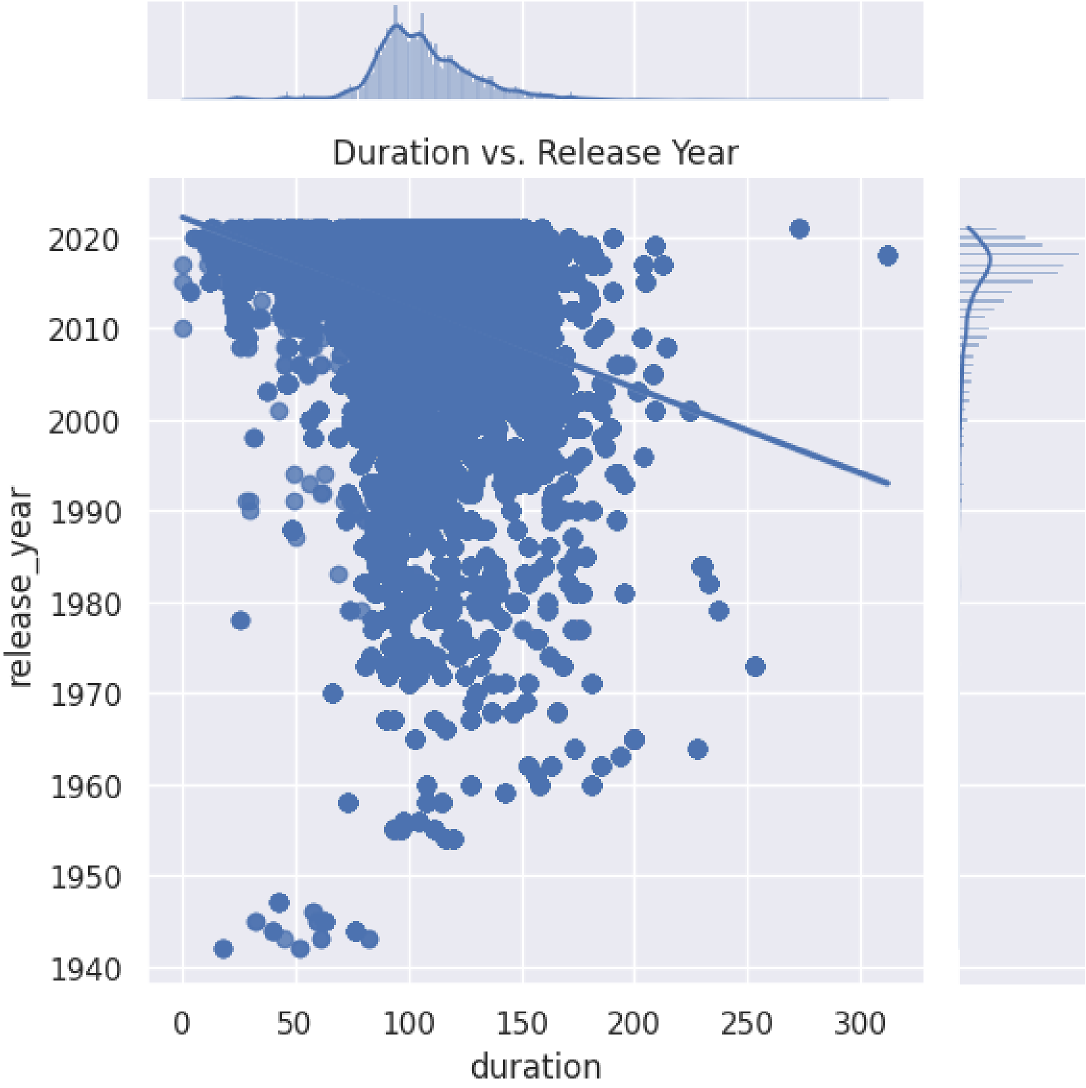
Insights :as per the heat map , in most of the countries, international movies genre is the most popular genre .Documentaries and Dramas are also in increasing trend across the countries. Categories like Comdies, musicals, romatic movies have lesser takes across the globe.

Recommendations : More of international movies could source as part of the global uptake. But this map could also be used to cater the needs of the regional crowd too. The interests could be seen individually on coutnry basis and those categories of movies could be added to the platform accordingly.

Q15) What is the distribution of duration against various release years ?



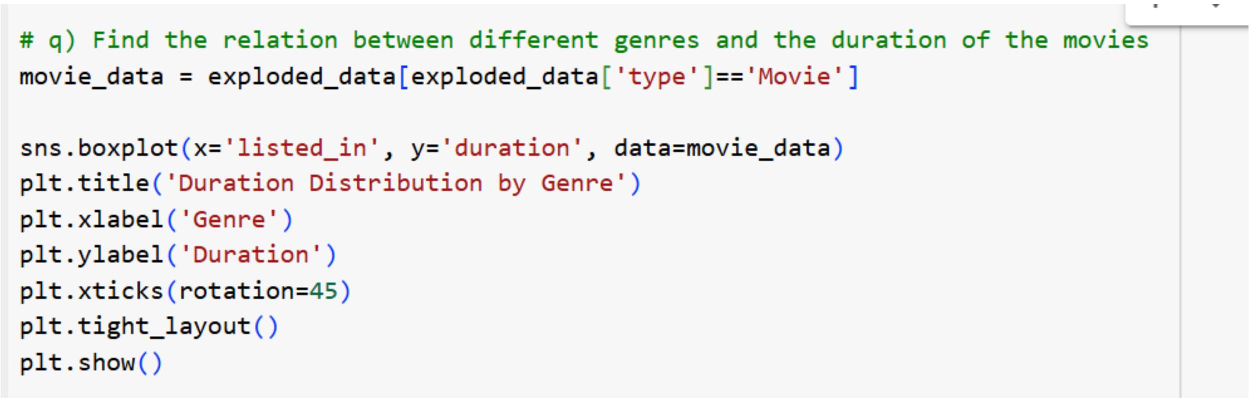
```
# Q) Distribution of duration against various release year

movie_data = exploded_data[exploded_data['type']=='Movie']
sns.jointplot(x='duration', y='release_year', data=movie_data, kind='reg')
plt.title('Duration vs. Release Year')
plt.tight_layout()
plt.show()
```



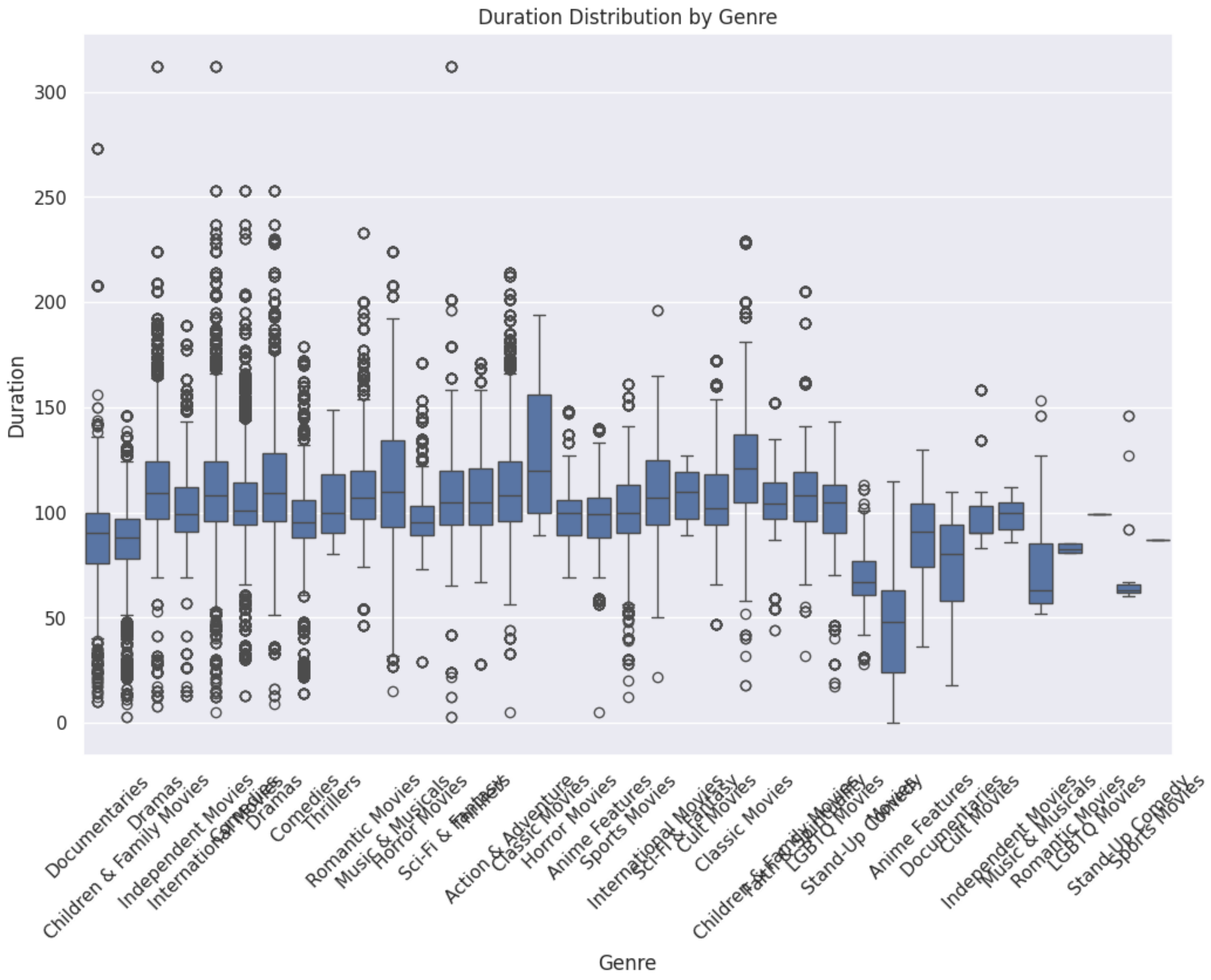
Insights : it should be noted that the duration is between 100 and 150 mins that are being most likely watched by the users in the recent timings. This means that they are more interested in medium range of time . Less than 2 hours to time.

Recommendations : More movies that comes in the range of 2 hours or less could be sourced to the platform. It should also be noted that there are takers for other duration too. So we need to just focus on less than 2 hours content. If the content is good, the people would be ready to watch for more than 2 hours.

Q16)

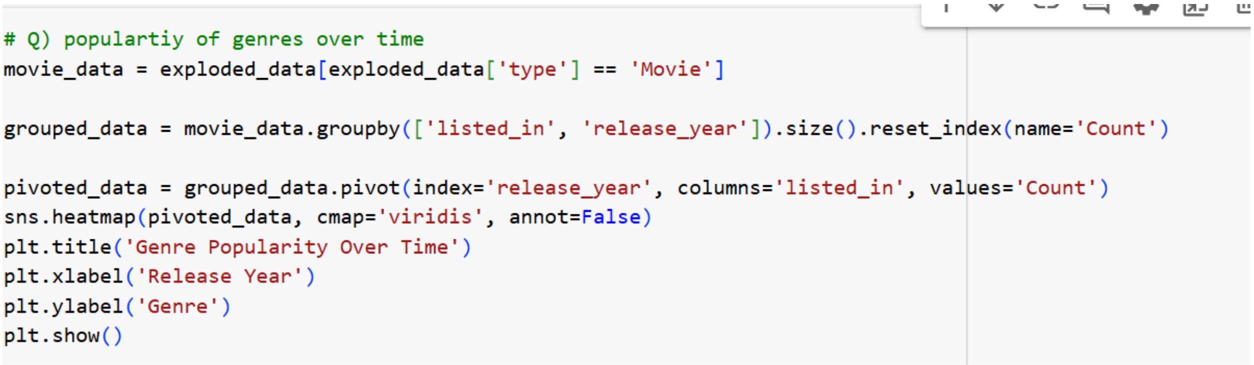


```
# q) Find the relation between different genres and the duration of the movies
movie_data = exploded_data[exploded_data['type']=='Movie']

sns.boxplot(x='listed_in', y='duration', data=movie_data)
plt.title('Duration Distribution by Genre')
plt.xlabel('Genre')
plt.ylabel('Duration')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

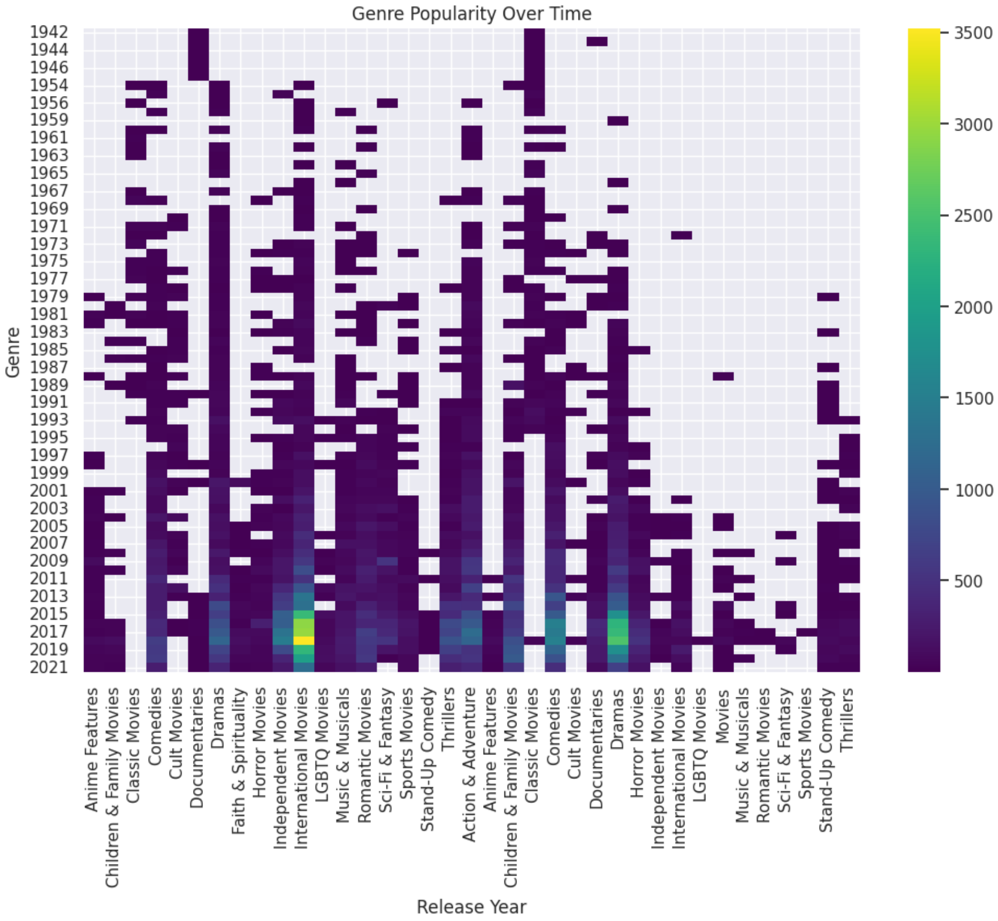


Insights : Genre Variation: The box plot reveals that the duration distribution varies significantly across different genres. Some genres, like Documentaries and Children & Family Movies, tend to have shorter durations, while others, such as Action & Adventure and Sci-Fi & Fantasy, have longer durations.
Overlapping Distributions: There is some overlap between the distributions of certain genres, suggesting that there is not a strict correlation between genre and duration.
Outliers: The presence of outliers indicates that some movies within certain genres have significantly different durations compared to the majority.

Q17) Popularity of genres over time.



```
# Q) populartiy of genres over time
movie_data = exploded_data[exploded_data['type'] == 'Movie']

grouped_data = movie_data.groupby(['listed_in', 'release_year']).size().reset_index(name='Count')

pivoted_data = grouped_data.pivot(index='release_year', columns='listed_in', values='Count')
sns.heatmap(pivoted_data, cmap='viridis', annot=False)
plt.title('Genre Popularity Over Time')
plt.xlabel('Release Year')
plt.ylabel('Genre')
plt.show()
```

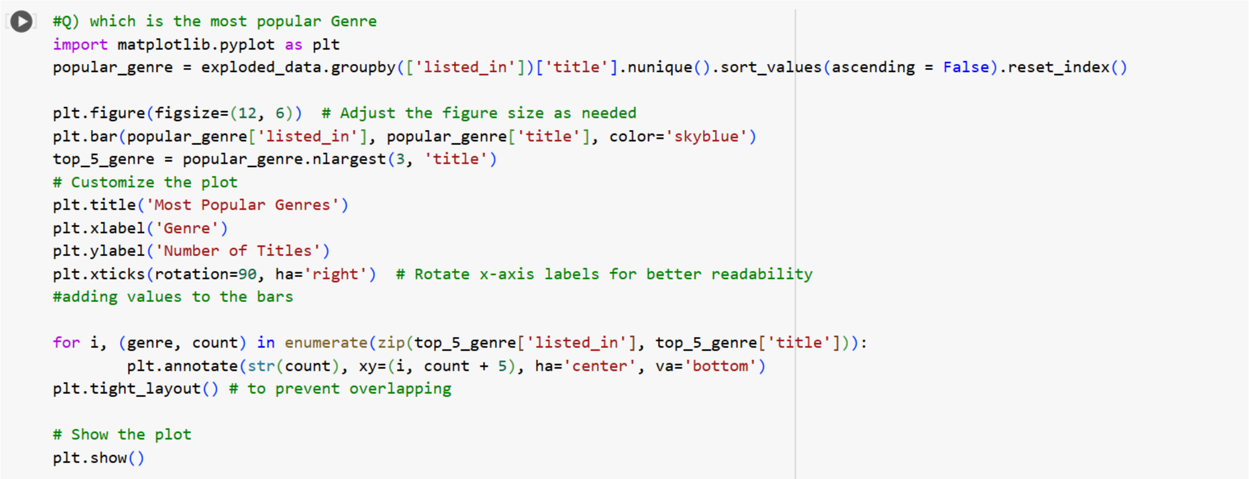


Insights: in the recent years international movies are the most popular genres along with Drama. It should be noted that these genres are not a new one, but they have been under watch for a long period of time and it was in the recent years that their viewership has been increased. This could be due to increase in availability of the content via various OTT platforms and also the increase in internet penetration globally which has improved the access of global data locally.

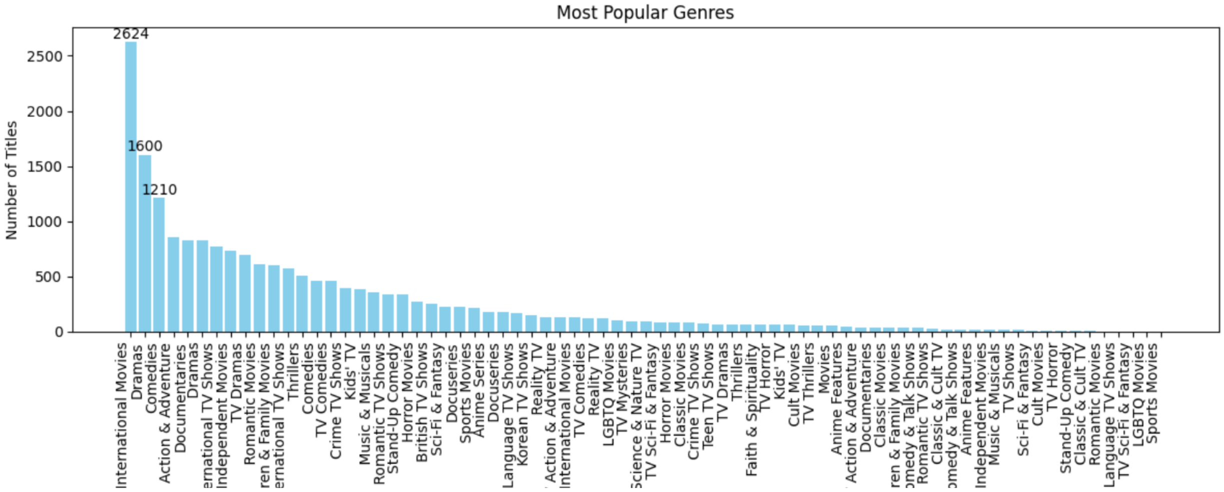
Stand up comedy, anime are some of the other groups which was started in the recent years and gaining traction.

Recommendations : More on International movies could be added along with other categories of Drama, Standup comedy anime, etc. This can also be plotted for various regions separately and then add contents accordingly which would be a more targeted approach.

Q18)



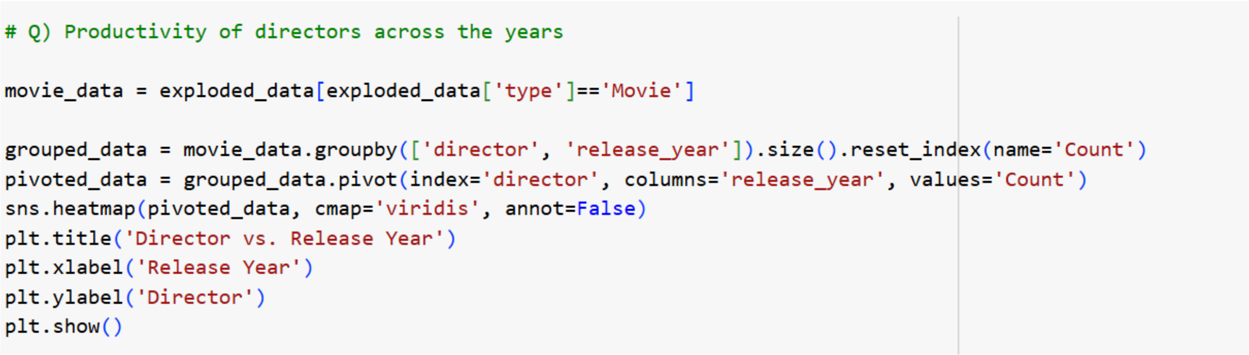
```
#Q) which is the most popular Genre
import matplotlib.pyplot as plt
popular_genre = exploded_data.groupby(['listed_in'])['title'].nunique().sort_values(ascending = False).reset_index()

plt.figure(figsize=(12, 6))  # Adjust the figure size as needed
plt.bar(popular_genre['listed_in'], popular_genre['title'], color='skyblue')
top_5_genre = popular_genre.nlargest(3, 'title')
# Customize the plot
plt.title('Most Popular Genres')
plt.xlabel('Genre')
plt.ylabel('Number of Titles')
plt.xticks(rotation=90, ha='right')  # Rotate x-axis labels for better readability
#adding values to the bars

for i, (genre, count) in enumerate(zip(top_5_genre['listed_in'], top_5_genre['title'])):
        plt.annotate(str(count), xy=(i, count + 5), ha='center', va='bottom')
plt.tight_layout() # to prevent overlapping

# Show the plot
plt.show()
```



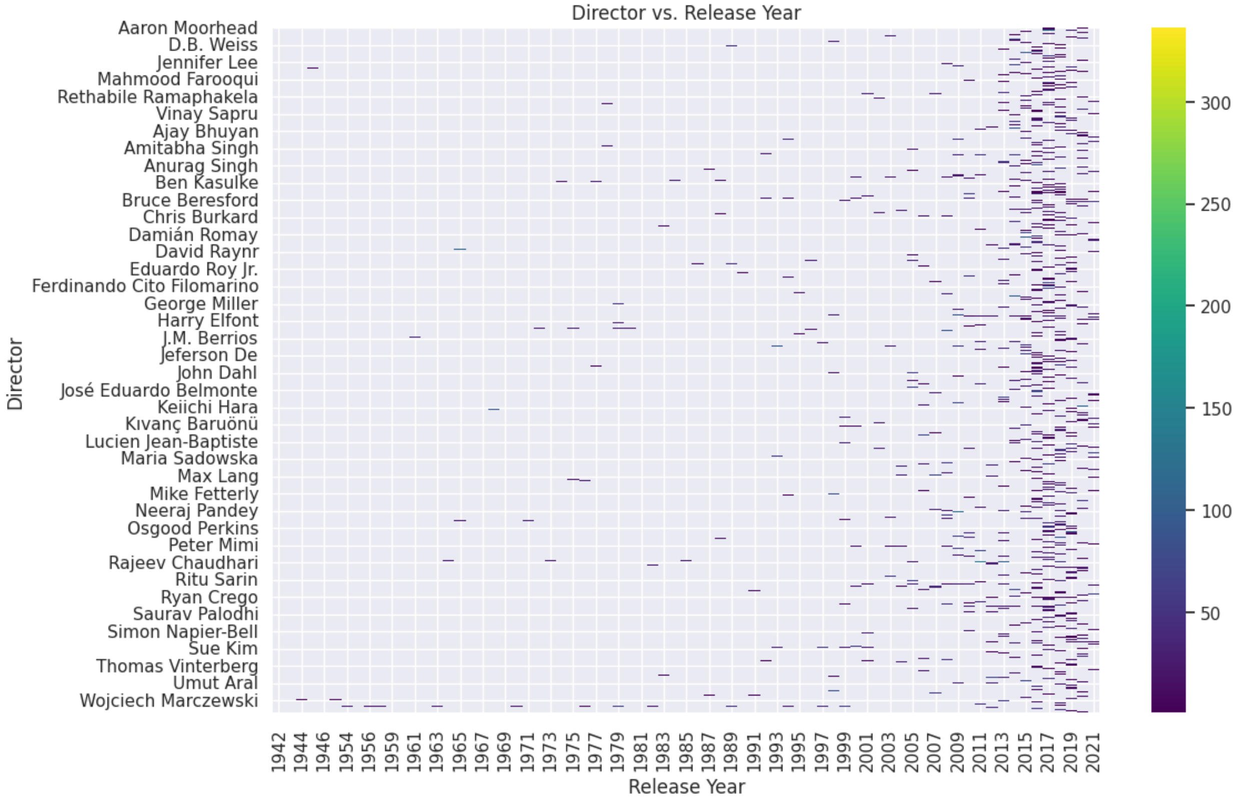
insights : International movies with 2624 count is the most popular genre . The second popular is the Dramas section with 1600 and the third popular genre is Comedies with 1210.

Recommendations : this could be a leading light for getting more contents when selecting to be added to the platform.

Q19)

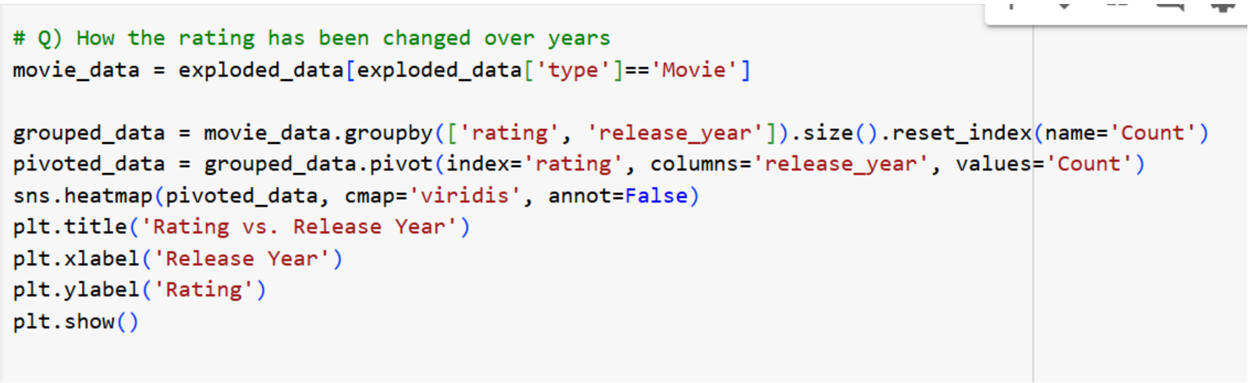


```
# Q) Productivity of directors across the years

movie_data = exploded_data[exploded_data['type']=='Movie']

grouped_data = movie_data.groupby(['director', 'release_year']).size().reset_index(name='Count')
pivoted_data = grouped_data.pivot(index='director', columns='release_year', values='Count')
sns.heatmap(pivoted_data, cmap='viridis', annot=False)
plt.title('Director vs. Release Year')
plt.xlabel('Release Year')
plt.ylabel('Director')
plt.show()
```

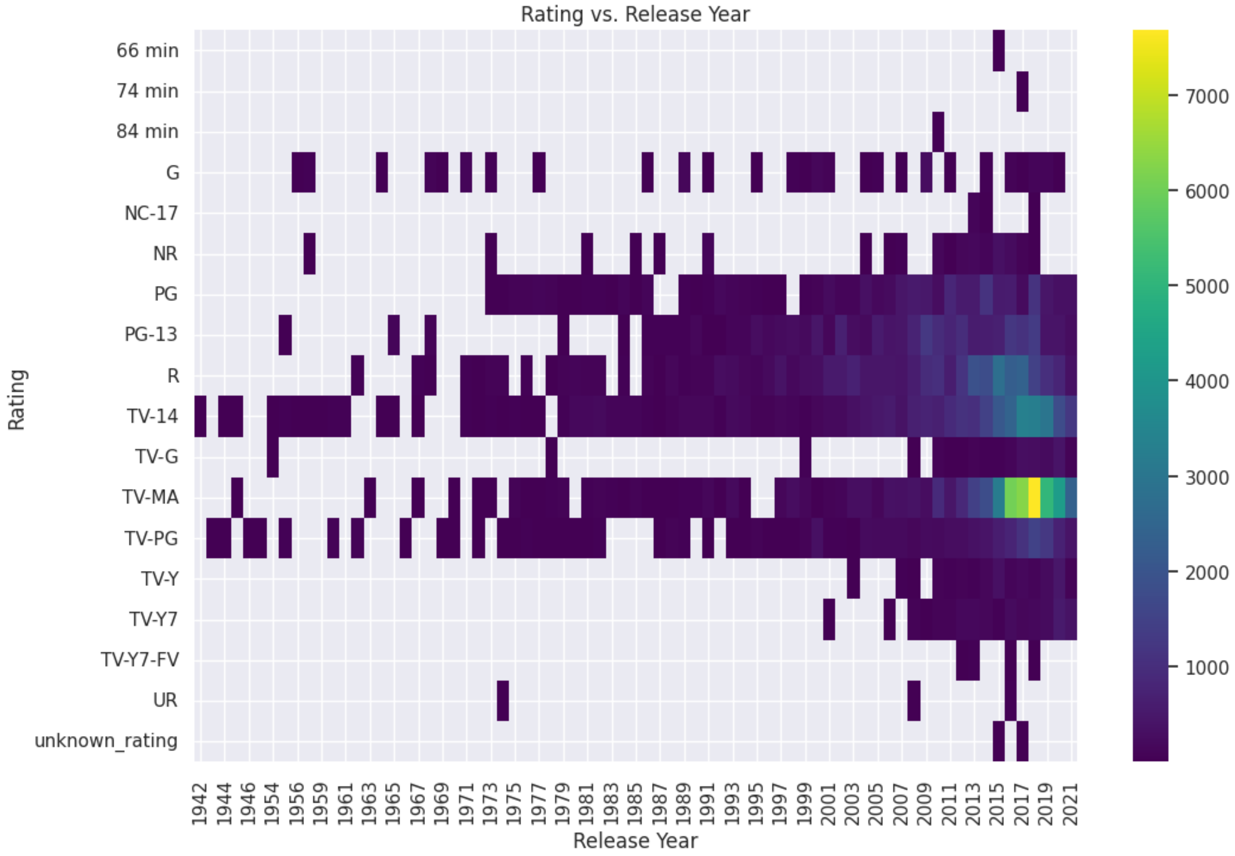


Insights: this says that the directors are most productive in the recent years. This could be attributed to the recent development in technology that has helped them to direct movies in a better position with all the accessories they would require for the better visualisation.

Q20)

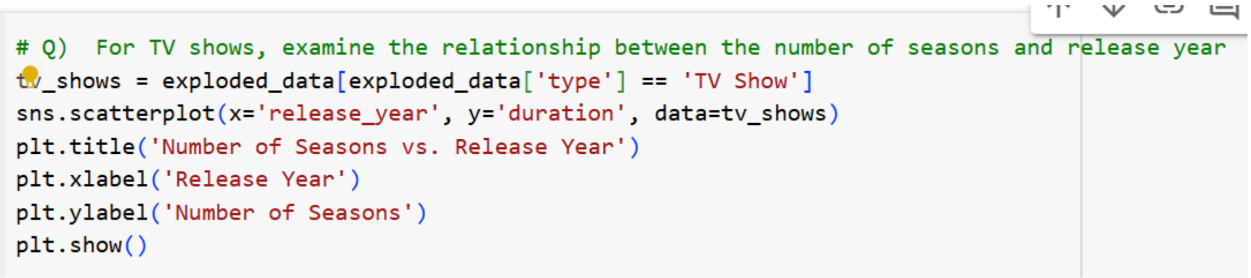


```
# Q) How the rating has been changed over years
movie_data = exploded_data[exploded_data['type']=='Movie']

grouped_data = movie_data.groupby(['rating', 'release_year']).size().reset_index(name='Count')
pivoted_data = grouped_data.pivot(index='rating', columns='release_year', values='Count')
sns.heatmap(pivoted_data, cmap='viridis', annot=False)
plt.title('Rating vs. Release Year')
plt.xlabel('Release Year')
plt.ylabel('Rating')
plt.show()
```

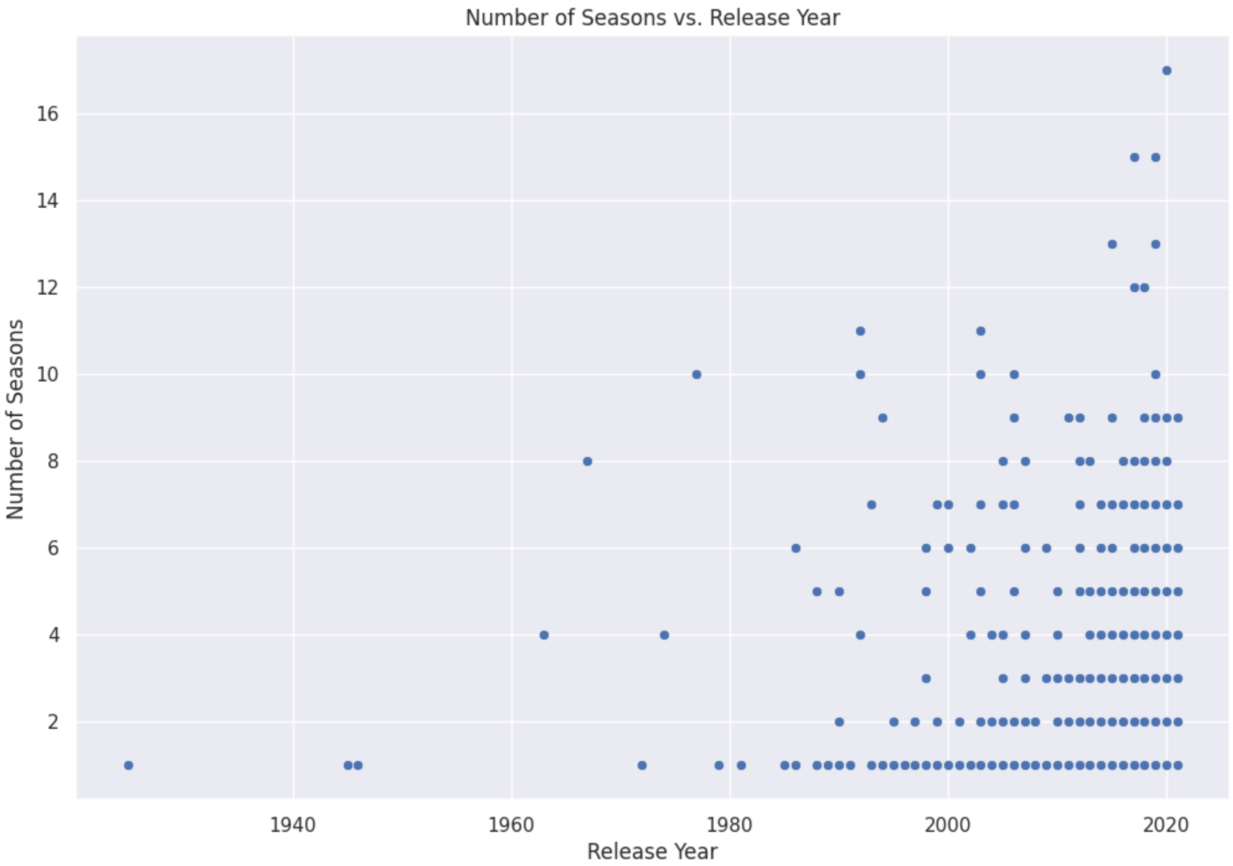


Insights : TV_MA section has been increased in the platform in the recent years. Along with that TV-14 rating is also increasing. The next best rating would be TV-PG. The increase in rating for these sections could be attributed to the boom in internet penetration and the increase in the smart phone accessibility.This is also accompanied with the digital conversion of the contents and their availabilty in the platforms with afforadable plans for subscription.

Q21)



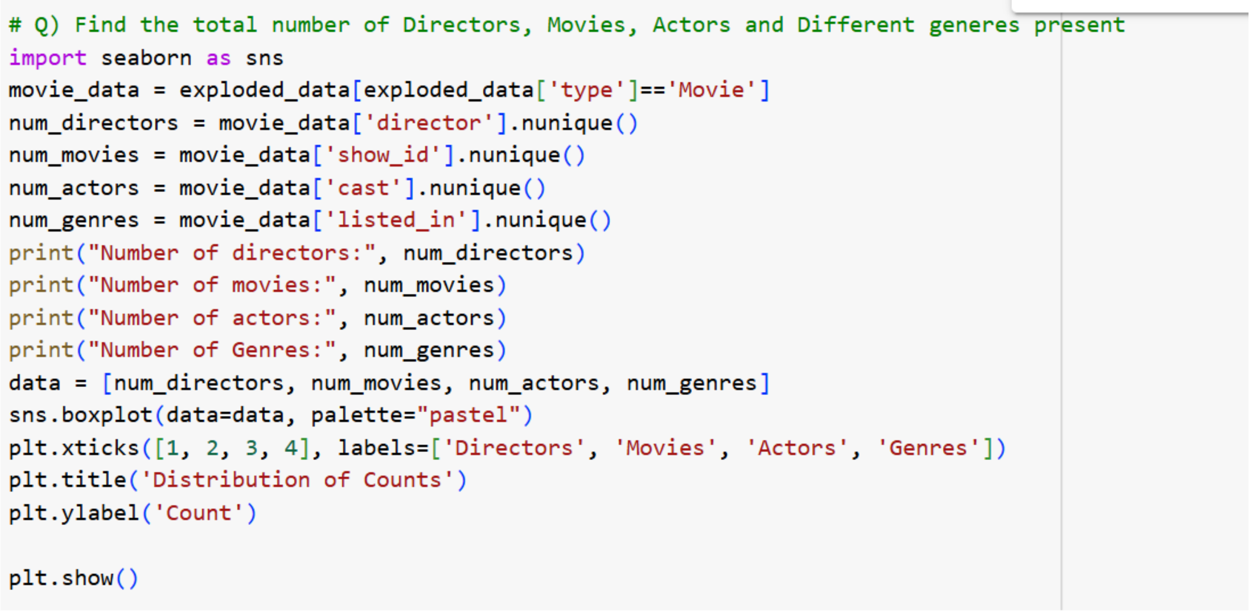
```
# Q)  For TV shows, examine the relationship between the number of seasons and release year
tv_shows = exploded_data[exploded_data['type'] == 'TV Show']
sns.scatterplot(x='release_year', y='duration', data=tv_shows)
plt.title('Number of Seasons vs. Release Year')
plt.xlabel('Release Year')
plt.ylabel('Number of Seasons')
plt.show()
```



Insights : Most of the tv shows are having one seasons and those has been increased in the recent years. It must also be The other shows with more number of seasons are also in the increasing trend. The average number of seasons for the TV shows present was found as 8. so this tells that people are interested in seeing contents which are long as well as short.

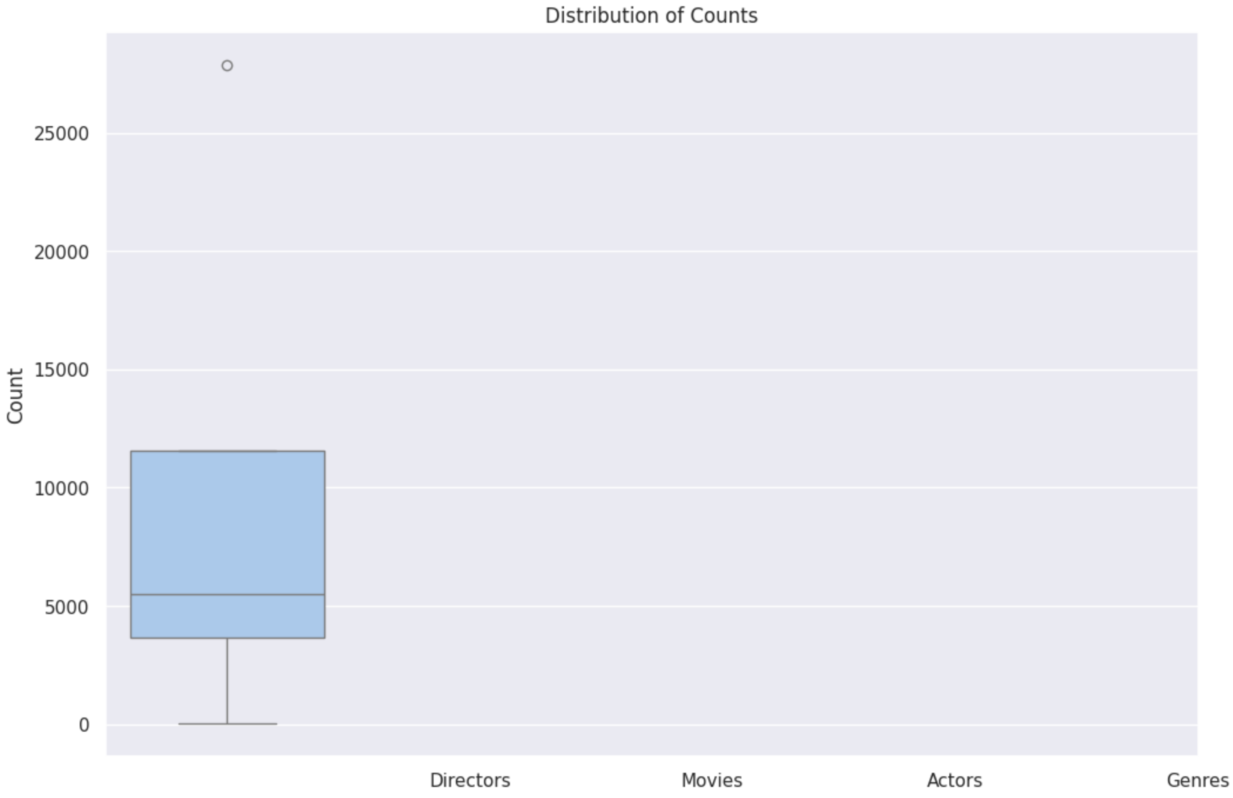
Recommendations : Platform should take care of creating contents that are short as well as long format. But it should also be noted that not just the number of seasons are important, the viewers should be able to co relate themselves with the characters present in the shows so that they also move along with the increasing season count. Content quality should be of prime focus than the quantity. If a good content is shot over various seasons then it can be counted in.

Q22)



```
# Q) Find the total number of Directors, Movies, Actors and Different generes present
import seaborn as sns
movie_data = exploded_data[exploded_data['type']=='Movie']
num_directors = movie_data['director'].nunique()
num_movies = movie_data['show_id'].nunique()
num_actors = movie_data['cast'].nunique()
num_genres = movie_data['listed_in'].nunique()
print("Number of directors:", num_directors)
print("Number of movies:", num_movies)
print("Number of actors:", num_actors)
print("Number of Genres:", num_genres)
data = [num_directors, num_movies, num_actors, num_genres]
sns.boxplot(data=data, palette="pastel")
plt.xticks([1, 2, 3, 4], labels=['Directors', 'Movies', 'Actors', 'Genres'])
plt.title('Distribution of Counts')
plt.ylabel('Count')

plt.show()
```

```
Number of directors: 4887
Number of movies: 6131
Number of actors: 27880
Number of Genres: 37
```



Insights : The total number of directors are 4887 over all the countries and languages. There are 27880 actors who have acted in 6131 movies that were released and added to the platform.

Recommendations : with the increase in technology help, the directors are able to add more movies to their career and thus employ more number of actors for the same. The increase in number of directors, actors and movies indicates that movies will always be a sweet spot for the talented and passionate ones.