

Mindset

1. Evaluation will be kept lenient, so make sure you attempt this case study.
2. Read the question carefully and try to understand what exactly is being asked.
3. Brainstorm a little. If you're getting an error, remember that Google is your best friend.
4. You can watch the lecture recordings or go through your lecture notes once again if you feel like you're getting confused over some specific topics.
5. Discuss your problems with your peers. Make use of the Slack channel and WhatsApp group.
6. Only if you think that there's a major issue, you can reach out to your Instructor via Slack or Email.
7. There is no right or wrong answer. We have to get used to dealing with uncertainty in business. This is exactly the skill we want to develop.

About NETFLIX

Netflix is one of the most popular media and video streaming platforms. They have over 10000 movies or tv shows available on their platform, as of mid-2021, they have over 222M Subscribers globally. This tabular dataset consists of listings of all the movies and tv shows available on Netflix, along with details such as - cast, directors, ratings, release year, duration, etc.

Business Problem

Analyze the data and generate insights that could help Netflix in deciding which type of shows/movies to produce and how they can grow the business in different countries

Dataset

Link: [Dataset link](#)

(After clicking on the above link, you can download the files by right-clicking on the page and clicking on "Save As", then naming the file as per your wish, with .csv as the extension.)

The dataset provided to you consists of a list of all the TV shows/movies available on Netflix:

Show_id: Unique ID for every Movie / Tv Show

Type: Identifier - A Movie or TV Show

Title: Title of the Movie / Tv Show

Director: Director of the Movie

Cast: Actors involved in the movie/show

Country: Country where the movie/show was produced

Date_added: Date it was added on Netflix

Release_year: Actual Release year of the movie/show

Rating: TV Rating of the movie/show

Duration: Total Duration - in minutes or number of seasons

Listed_in: Genre

Description: The summary description

Hints

1. The exploration should have a goal. As you explore the data, keep in mind that you want to answer which type of shows to produce and how to grow the business.
2. Ensure each recommendation is backed by data. The company is looking for data-driven insights, not personal opinions or anecdotes.
3. Assume that you are presenting your findings to business executives who have only a basic understanding of data science. Avoid unnecessary technical jargon.
4. Start by exploring a few questions: What type of content is available in different countries?
 1. How has the number of movies released per year changed over the last 20-30 years?
 2. Comparison of tv shows vs. movies.
 3. What is the best time to launch a TV show?
 4. Analysis of actors/directors of different types of shows/movies.
 5. Does Netflix has more focus on TV Shows than movies in recent years
 6. Understanding what content is available in different countries

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

netflix = pd.read_csv("Netflix_data.csv")

print(netflix.dtypes)
```

```
show_id      object
type         object
title        object
director     object
cast         object
country      object
date_added   object
release_year  int64
rating       object
duration     object
listed_in    object
description   object
dtype: object
```

```
print(netflix.shape)
```

```
(8807, 12)
```

Missing Value Check :

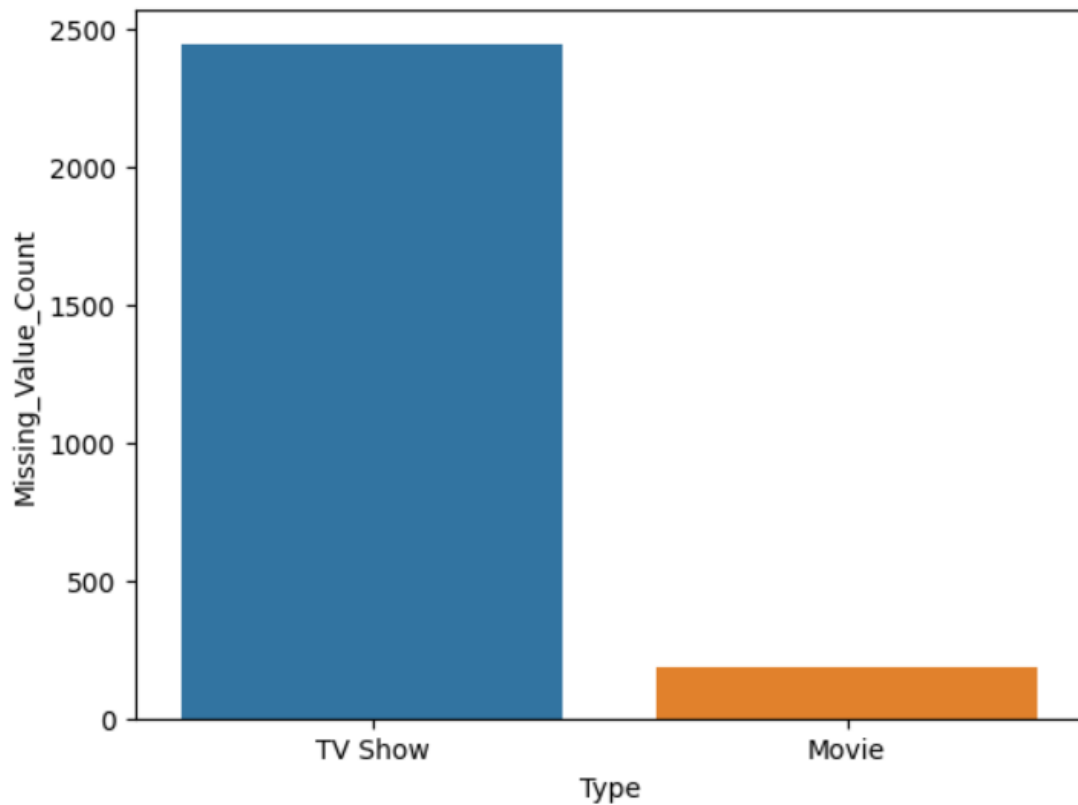
```
print(netflix.isnull().sum())
```

```
show_id      0
type         0
title        0
director     2634
cast         825
country      831
date_added   10
release_year  0
rating       4
duration     3
listed_in    0
description   0
dtype: int64
```

```
dire_missing_value = netflix.loc[netflix["director"].isna()]
A = dire_missing_value.groupby("type")["title"].count()
print(A)
```

```
Movie      188
TV Show    2446
```

```
sns.countplot(x = "type" , data = dire_missing_value)
plt.xlabel("Type")
plt.ylabel("Missing_Value_Count")
```



```
print(netflix.describe())
```

	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

Observation :

1. The shape of the dataset is (8807,12). Which means the data has 8807 rows and 12 columns.

2. All the columns are of "Object" datatype except release_year column("int").

3. Missing values are present in director, cast, country and date_added column. for the calculation I am going to ignore the missing values in

cast, country and date_added column as they are very less in numbers. and maximum missing values are present for TV Show in director column.

4. As there is only one numeric value present in the dataset which is release_year it's statistical summary is printed above.

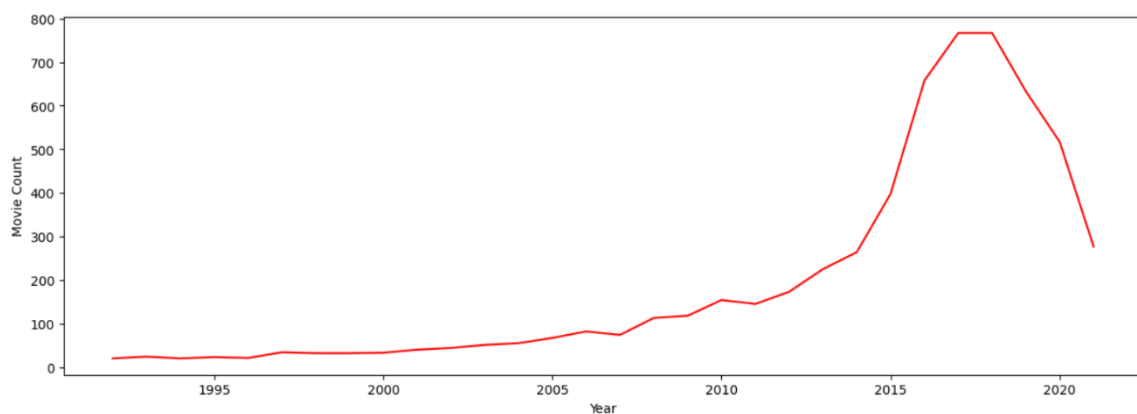
1.) How has the number of movies released per year changed over the last 20-30 years?

```
netflix_shape = netflix.shape
netflix_movie_count = netflix.loc[netflix["type"] ==
"Movie" , "release_year"].value_counts().sort_index(ascending
= False).head(30)

x_bar = netflix_movie_count.index
y_bar = netflix_movie_count
plt.figure(figsize = (15 ,5))
sns.lineplot(x= x_bar , y = y_bar , data = netflix_movie_count
, color = "r")
plt.xlabel("Year")
plt.ylabel("Movie Count")
plt.show()
```

OUTPUT :

```
2021    277
2020    517
2019    633
2018    767
2017    767
2016    658
2015    398
2014    264
2013    225
2012    173
2011    145
2010    154
2009    118
2008    113
2007     74
2006     82
2005     67
2004     55
2003     51
2002     44
2001     40
2000     33
1999     32
1998     32
1997     34
1996     21
1995     23
1994     20
1993     24
1992     20
Name: release_year, dtype: int64
```



Observation :

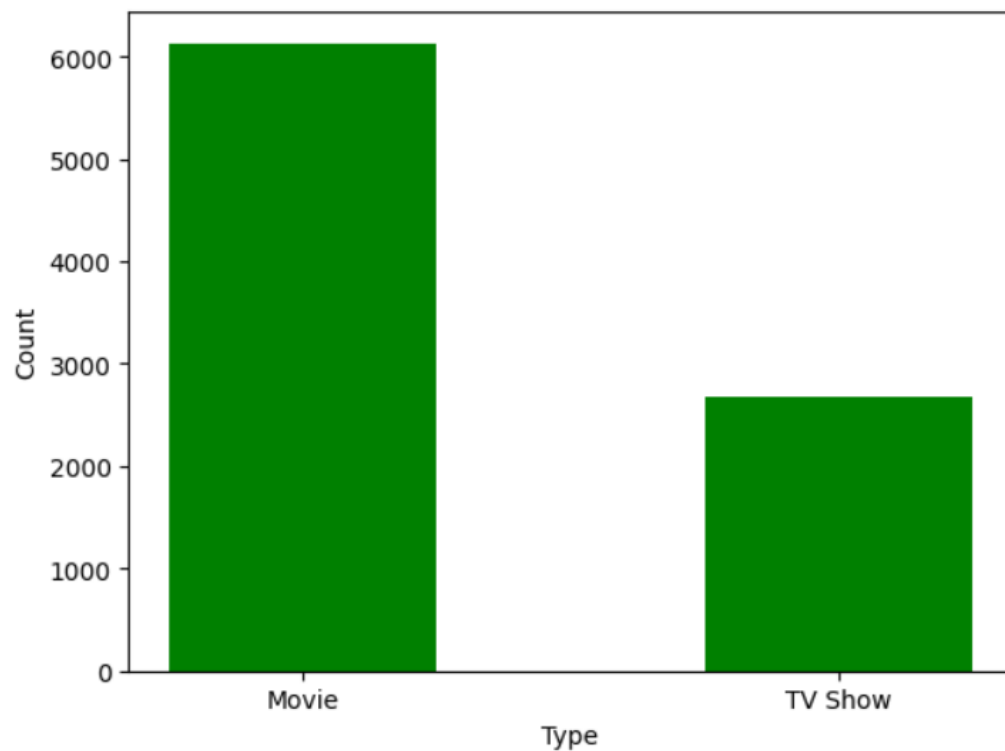
1. As we can see from the output that the Netflix data has 8807 rows and 12 columns.
2. We have most missing values in "director" , "cast" and "country" column as per dataframe information.
3. As we can see from the graph and output that the movies that were released between 2015 and 2021 are more likely to be on Netflix.
4. Netflix has less old released movies than new released movies on platform.
5. Netflix has more movies that were released from 2015 to 2020 but there is sudden drop in the movie release in 2021 and
hence less movie addition on platform this is because of the pandemic.

2.) Comparison of tv shows vs. movies.

```
A = netflix.type.value_counts()
print(A)
plt.bar(A.index , A , color = "green" , width = 0.5)
plt.xlabel("Type")
plt.ylabel("Count")
plt.show()
```

Output :

```
Movie      6131
TV Show    2676
Name: type, dtype: int64
```



```
net = netflix.copy()
type_movie = net[net["type"] == "Movie"]
type_show = net[net["type"] == "TV Show"]
type_movie["duration"].fillna(0 , inplace = True)
type_show["duration"].fillna(0 , inplace = True)

type_movie = type_movie["duration"].str.replace(" min" , "")
type_movie = type_movie.astype({ "duration" : "float"})
type_movie = type_movie.mean()

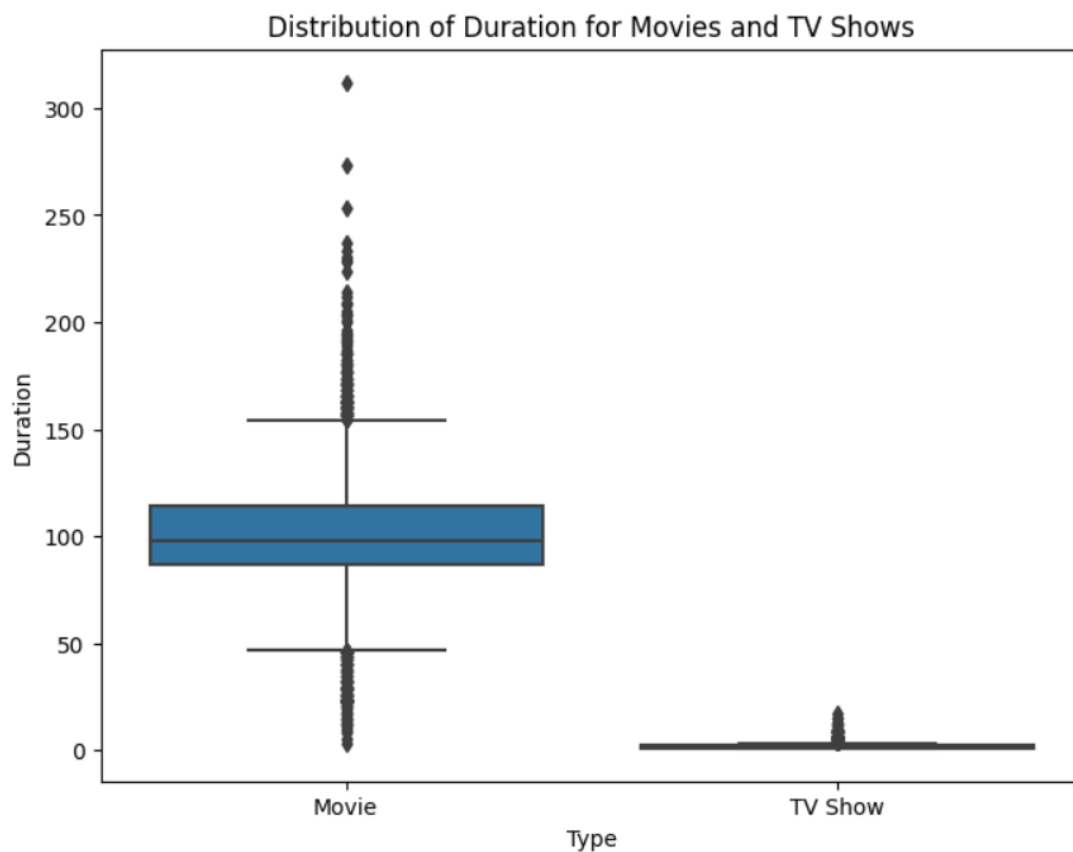
type_show = type_show["duration"].str.split(" " , expand =
True)
type_show = type_show.T.rename({0
:"value"}).T.reset_index(drop = True)
type_show = type_show.astype({ "value" : "float"})
type_show = type_show["value"].mean()

print(f"Movie_Mean_Duration : {type_movie}")
print(f"Show_Mean_Duration : {type_show}")
```

Output :


```
Movie_Mean_Duration : 99.57718668407311
Show_Mean_Duration : 1.764947683109118
```

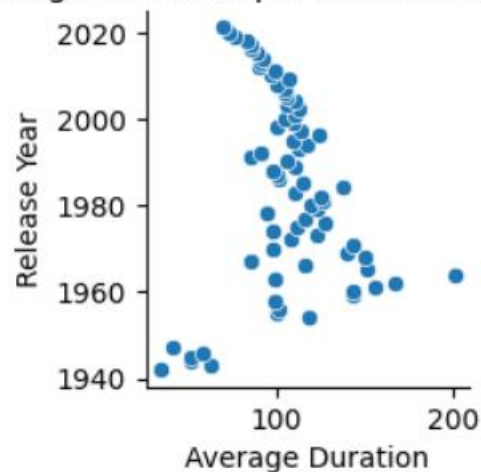
```
net['duration'] =
net['duration'].str.extract('(\d+)').astype(float)
plt.figure(figsize=(8, 6))
sns.boxplot(x='type', y='duration', data=net)
plt.title("Distribution of Duration for Movies and TV Shows")
plt.xlabel("Type")
plt.ylabel("Duration")
plt.show()
```



```
netflix['duration'] = netflix['duration'].replace('1 Season',
1)
netflix['duration'] =
netflix['duration'].str.extract('(\d+)').astype(float)
average_duration_per_year =
netflix.groupby('release_year')['duration'].mean().reset_index
()
```

```
plt.figure(figsize=(10, 8))
sns.pairplot(data=average_duration_per_year,
x_vars='duration', y_vars='release_year', kind='scatter')
plt.title("Pairplot: Average Duration per Release Year vs.
Release Year")
plt.xlabel("Average Duration")
plt.ylabel("Release Year")
plt.show()
```

Pairplot: Average Duration per Release Year vs. Release Year



Observation :

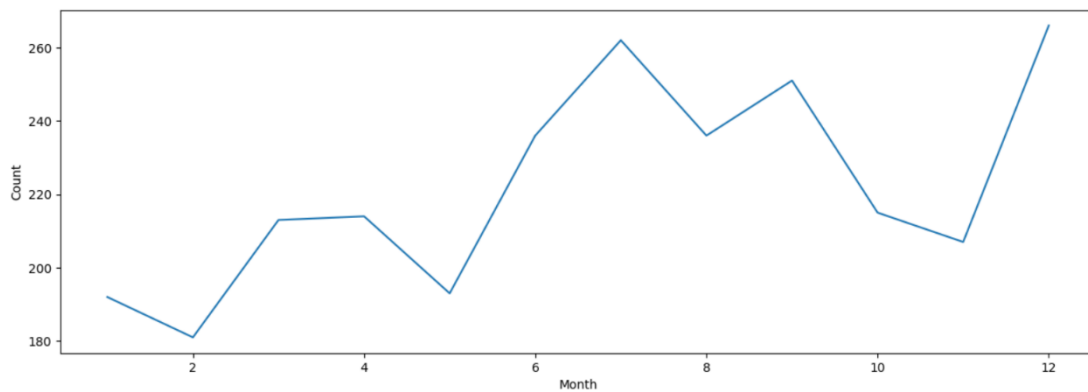
1. As we can see from the graph and the data that Netflix has movies two times than TV Show.
2. We can conclude from this is that people like to watch more movies than a TV Show.
3. One reason could be that Movies are having average duration of ~100 minutes and TV Shows are having ~2 seasons, due to more time to complete a TV Show than Movies people prefer to watch Movies over Shows.

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

```
A = netflix.loc[netflix["type"] == "TV Show",  
               "month"].value_counts()  
print(A)  
plt.figure(figsize = (15, 5))  
sns.lineplot(x = A.index , y = A , data = A)  
plt.xlabel("Month")  
plt.ylabel("Count")  
plt.show()
```

Output :

```
12.0    266  
7.0     262  
9.0     251  
8.0     236  
6.0     236  
10.0    215  
4.0     214  
3.0     213  
11.0    207  
5.0     193  
1.0     192  
2.0     181  
Name: month, dtype: int64
```



Observation :

1. As we can see from the graph and the output the best time to launch a TV Show is the 3rd quarter of the year which is

July, August, September and the best month is December as dec is the festival month in the world especially in the western region.

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

```
net = netflix.copy()
net_movie = netflix[netflix["type"] == "Movie"]
net_show = netflix[netflix["type"] == "TV Show"]

net_movie["director_nested"] =
net_movie["director"].str.split(", ")
net_movie = net_movie.explode("director_nested").reset_index()
net_movie_dire_count =
net_movie.groupby("director_nested")["title"].nunique().sort_v
alues(ascending = False).head(10)
print()
```

```

print(net_movie_dire_count)

net_movie["cast_nested"] = net_movie["cast"].str.split(", ")
net_movie = net_movie.explode("cast_nested").reset_index()
net_movie_cast_count =
net_movie.groupby("cast_nested")["title"].nunique().sort_values(
ascending = False).head(10)
print()
print(net_movie_cast_count)

plt.figure(figsize = (10 , 5))
plt.bar(net_movie_dire_count.index , net_movie_dire_count ,
color = "r")
plt.xticks(rotation = 90)
plt.xlabel("Movie_Director")
plt.ylabel("Movie_Count")

plt.figure(figsize = (10 , 5))
plt.bar(net_movie_cast_count.index , net_movie_cast_count ,
color = "g")
plt.xticks(rotation = 90)
plt.xlabel("Movie_Cast")
plt.ylabel("Movie_Count")

```

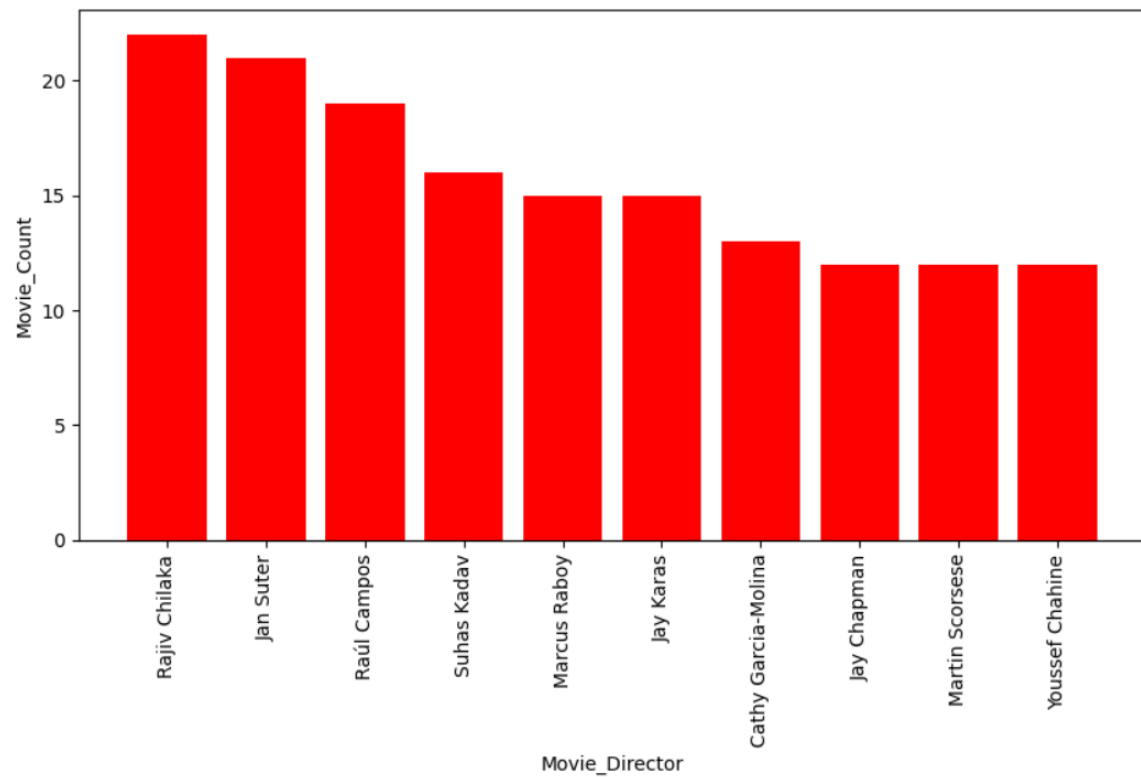
Output :

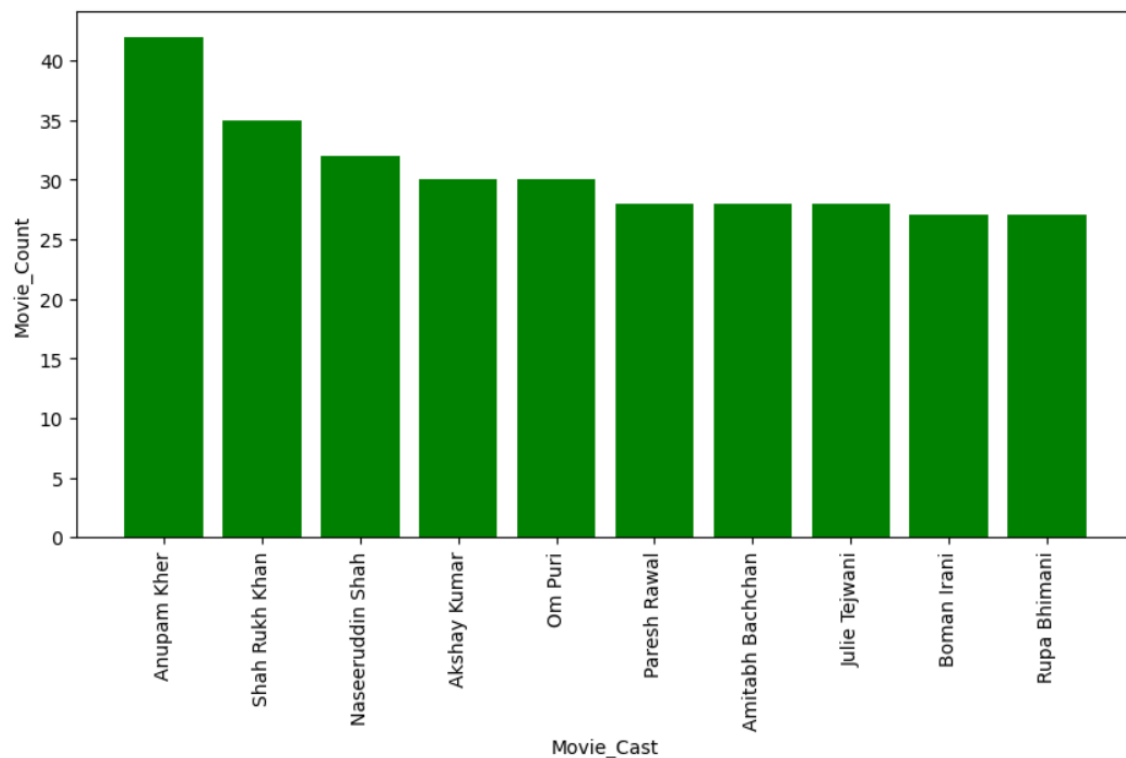
```

director_nested
Rajiv Chilaka      22
Jan Suter          21
Raúl Campos       19
Suhas Kadav       16
Marcus Raboy      15
Jay Karas         15
Cathy Garcia-Molina 13
Jay Chapman       12
Martin Scorsese   12
Youssef Chahine   12
Name: title, dtype: int64

```

```
cast_nested
Anupam Kher      42
Shah Rukh Khan   35
Naseeruddin Shah 32
Akshay Kumar     30
Om Puri          30
Paresh Rawal     28
Amitabh Bachchan 28
Julie Teiwani    28
Boman Irani      27
Rupa Bhimani     27
Name: title, dtype: int64
```





```
net_show["director_nested"] =  
net_show["director"].str.split(", ")  
net_show = net_show.explode("director_nested").reset_index()  
net_show_dire_count =  
net_show.groupby("director_nested")["title"].nunique().sort_val  
ues(ascending = False).head(10)  
print(net_show_dire_count)
```

```
net_show["cast_nested"] = net_show["cast"].str.split(", ")  
net_show = net_show.explode("cast_nested").reset_index()  
net_show_cast_count =  
net_show.groupby("cast_nested")["title"].nunique().sort_values  
(ascending = False).head(10)  
print(net_show_cast_count)
```

```

plt.figure(figsize = (10 , 5))
plt.bar(net_show_dire_count.index , net_show_dire_count ,
color = "b")
plt.xticks(rotation = 90)
plt.xlabel("Show_Director")
plt.ylabel("Show_Count")

plt.figure(figsize = (10 , 5))
plt.bar(net_show_cast_count.index , net_show_cast_count ,
color = "y")
plt.xticks(rotation = 90)
plt.xlabel("Show_Cast")
plt.ylabel("Show_Count")

```

Output :

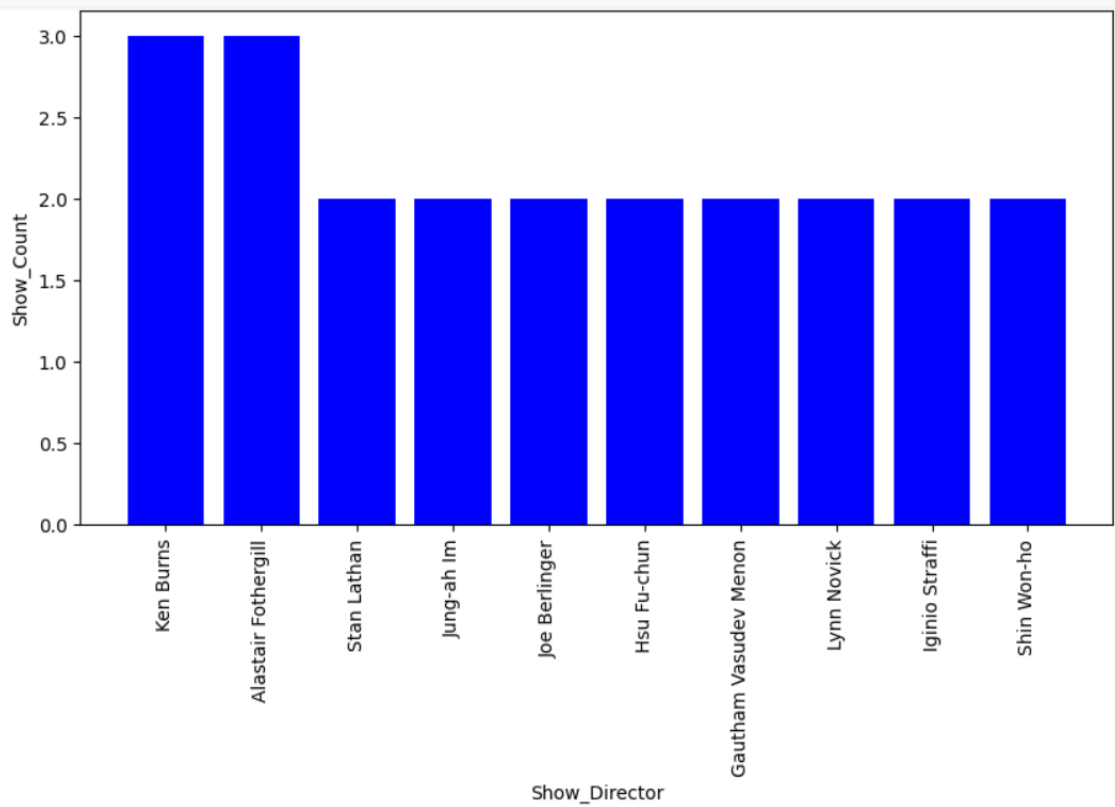
```

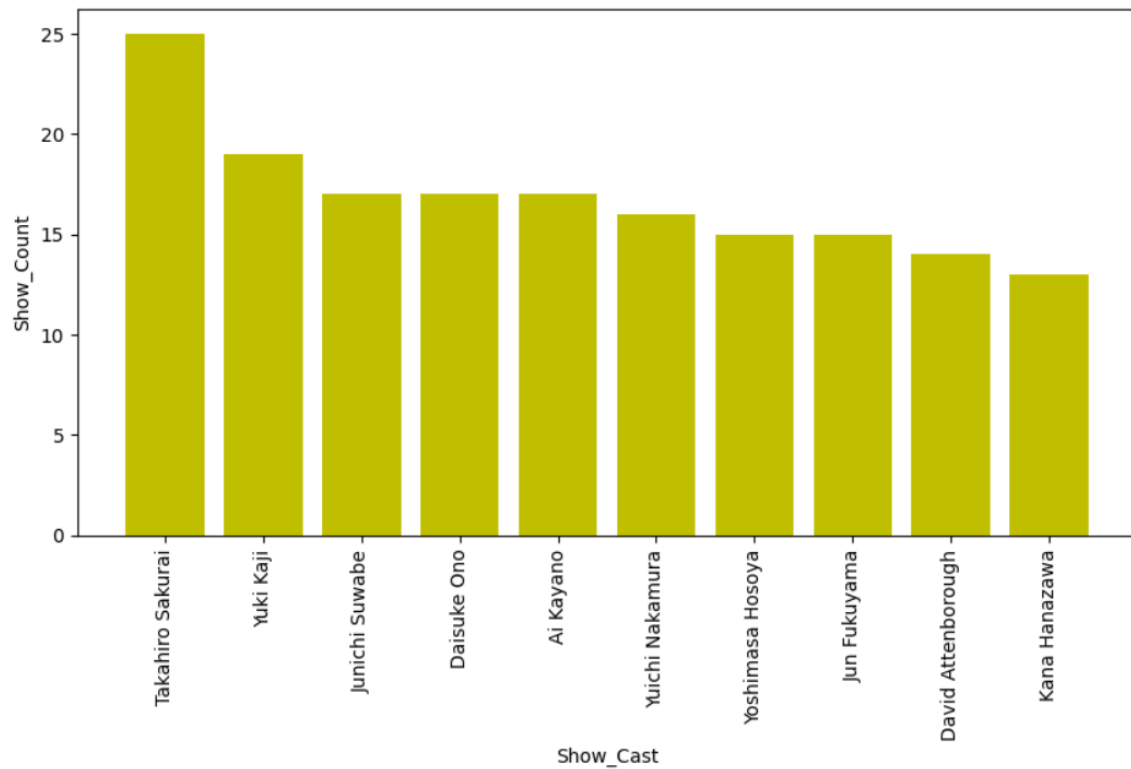
director_nested
Ken Burns      3
Alastair Fothergill  3
Stan Lathan    2
Jung-ah Im     2
Joe Berlinger  2
Hsu Fu-chun    2
Gautham Vasudev Menon  2
Lynn Novick    2
Iginio Straffi  2
Shin Won-ho    2
Name: title, dtype: int64

```



```
cast_nested
Takahiro Sakurai    25
Yuki Kaji           19
Junichi Suwabe      17
Daisuke Ono         17
Ai Kayano           17
Yuichi Nakamura     16
Yoshimasa Hosoya    15
Jun Fukuyama        15
David Attenborough  14
Kana Hanazawa       13
Name: title, dtype: int64
```





Observation :

1. Netflix has maximum movie directed by "Rajiv Chilaka" & "Anupam Kher" as an actor who has acted in maximum movies that is present on Netflix.
2. "Ken Burns" has directed maximum TV Shows & "Takahiro Sakurai" is an actor in most of the TV Shows.

5.) Does Netflix has more focus on TV Shows than movies in recent years.

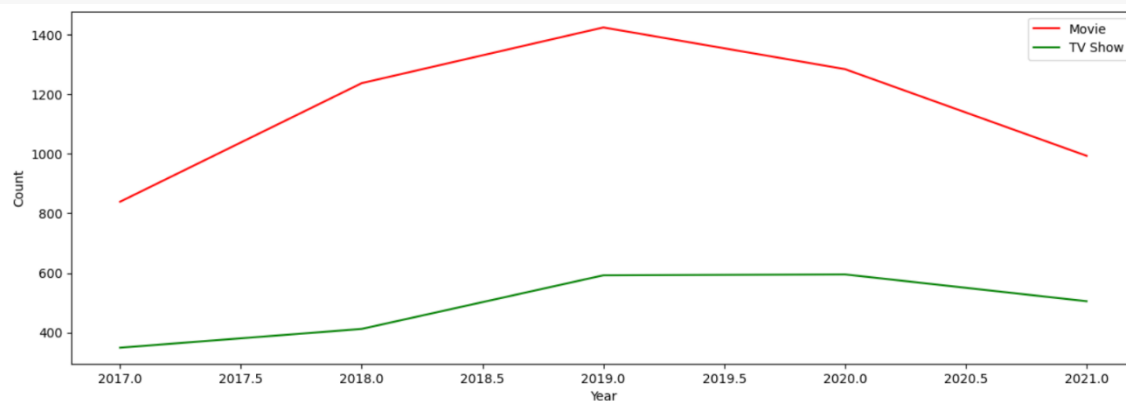
```
netflix["year"].fillna(0 , inplace = True)
netflix = netflix.astype({"year" : "int"})
max_year = netflix.year.max()
min_year = netflix.year.min()
print(max_year)
movie = netflix.loc[netflix["type"] == "Movie" ,
"year"].value_counts().sort_index().tail(5)
show = netflix.loc[netflix["type"] == "TV Show" ,
"year"].value_counts().sort_index().tail(5)
print(movie)
print()
print(show)

plt.figure(figsize = (15, 5))
sns.lineplot(x = movie.index , y = movie, data = movie , color
= "r" , label = "Movie")
sns.lineplot(x = show.index , y = show, data = show , color =
"g" , label = "TV Show")
plt.xlabel("Year")
plt.ylabel("Count")
plt.legend()
plt.show()
```

Output :

```
2017      839
2018     1237
2019     1424
2020     1284
2021      993
Name: year, dtype: int64
```

```
2017      349
2018      412
2019      592
2020      595
2021      505
Name: year, dtype: int64
```



Observation :

1. As we can see from the output in the recent years less movies got added on Netflix.
2. From plot we can say maximum movies got added on Netflix in 2019 and after 2019 there is a less addition of movies in every year.
3. There is no sudden fall in addition of TV Shows on Netflix after 2019. The graph for TV Show is almost static after 2019.
4. After comparing both Movies and TV Show plot we can say after 2019 less movies got added on Netflix as compare to TV Show with respect to their previous years.

6.) Understanding what content is available in different countries.

```
net_country = netflix.copy()
net_country["country_nested"] =
net_country["country"].str.split(", ")
net_country =
net_country.explode("country_nested").reset_index()

net_country["genre_nested"] =
net_country["listed_in"].str.split(", ")
net_country = net_country.explode("genre_nested").reset_index()

net_country = net_country.groupby(["country_nested" ,
"genre_nested"
])["title"].nunique().reset_index().sort_values("title" ,
ascending = False)
net_country_count =
net_country["country_nested"].value_counts().sort_values(ascending = False).head(10)

net_country_sum =
net_country.groupby("country_nested")["title"].sum().sort_values(ascending = False).head(10)

net_genre_count = net_country.groupby("genre_nested"
)["title"].nunique()
net_genre_count = net_genre_count.sort_values(ascending =
False).head(10)

def fun(df) :
    df["Total"] = df["title"].sum()
    return df
countrywise_genre =
net_country.groupby("country_nested").apply(fun).sort_values("
Total" , ascending = False)
countrywise_genre = countrywise_genre[["country_nested" ,
"genre_nested"]].head(30)

print(countrywise_genre)

print(net_country_count)
print()
print(net_country_sum)
print()
```

```
print(net_genre_count)

plt.figure(figsize = (15 , 5))
plt.bar(net_country_count.index , net_country_count , color =
"g")
plt.xlabel("Country")
plt.ylabel("Count_of_Unique_Genre")
plt.xticks(rotation = 90)
plt.show()

plt.figure(figsize = (15 , 5))
sns.countplot(x = "genre_nested" , data = net_country)
plt.xlabel("Genre")
plt.ylabel("Count_of_Genre")
plt.xticks(rotation = 90)
plt.show()
```

Output :

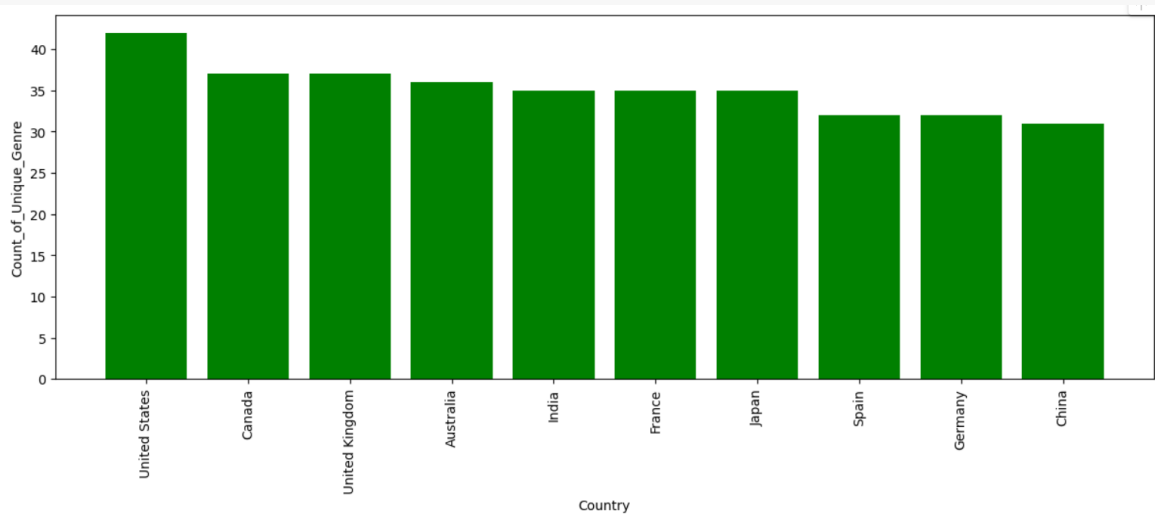
	country_nested	genre_nested
1364	United States	Romantic TV Shows
1340	United States	Anime Features
1361	United States	Music & Musicals
1347	United States	Crime TV Shows
1367	United States	Spanish-Language TV Shows
1362	United States	Reality TV
1379	United States	Teen TV Shows
1368	United States	Sports Movies
1371	United States	TV Action & Adventure
1374	United States	TV Horror
1352	United States	Faith & Spirituality
1351	United States	Dramas
1355	United States	International Movies
1377	United States	TV Shows
1358	United States	Korean TV Shows
1345	United States	Classic Movies
1366	United States	Science & Nature TV
1375	United States	TV Mysteries
1348	United States	Cult Movies
1376	United States	TV Sci-Fi & Fantasy
1356	United States	International TV Shows
1359	United States	LGBTQ Movies
1378	United States	TV Thrillers
1370	United States	Stand-Up Comedy & Talk Shows
1344	United States	Classic & Cult TV
1342	United States	British TV Shows
1373	United States	TV Dramas
1363	United States	Romantic Movies
1341	United States	Anime Series
1369	United States	Stand-Up Comedy

Unique Contents present in different countries:

```

United States    42
Canada          37
United Kingdom  37
Australia       36
India           35
France          35
Japan           35
Spain          32
Germany         32
China          31
Name: country_nested, dtype: int64

```



Count of genres appeared in different countries :

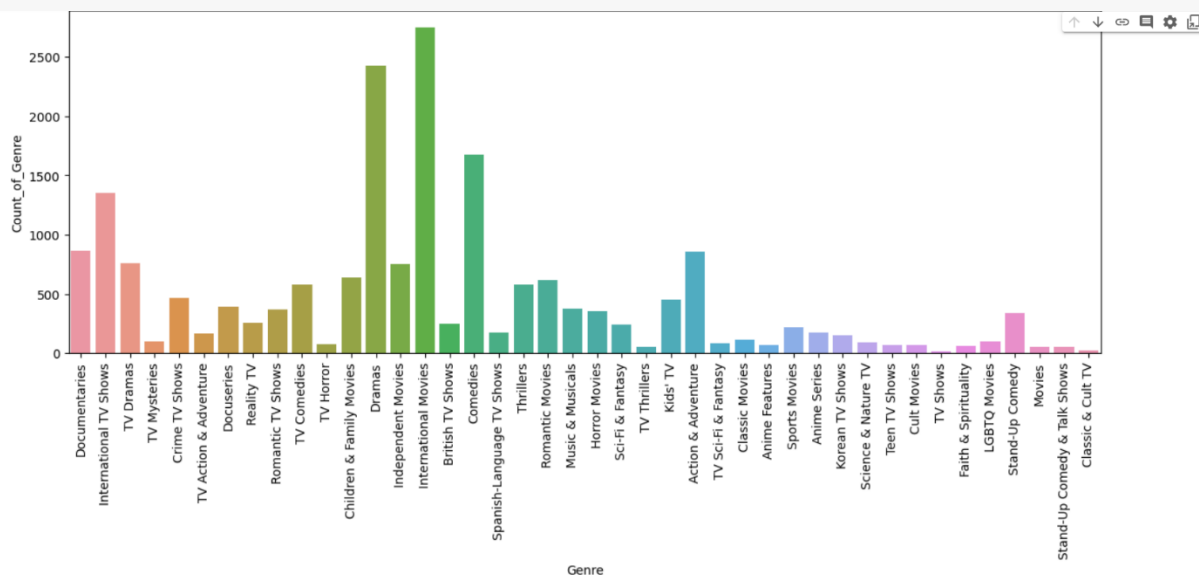
```

country_nested
United States    6779
India            2804
United Kingdom   1779
France           916
Canada           877
Japan            733
South Korea      632
Spain            591
Germany          511
Mexico           409
Name: title, dtype: int64

```


Count of different genres in the dataset :

	genre_nested	title
0	International Movies	2752
1	Dramas	2427
2	Comedies	1674
3	International TV Shows	1351
4	Documentaries	869
5	Action & Adventure	859
6	TV Dramas	763
7	Independent Movies	756
8	Children & Family Movies	641
9	Romantic Movies	616



```
A = len(net_country_count[net_country_count <= 10])
B = len(net_country_count)
print(f"Country with less geners available : {A}")
print(f"Length of all the countries : {B}")
```

Output :

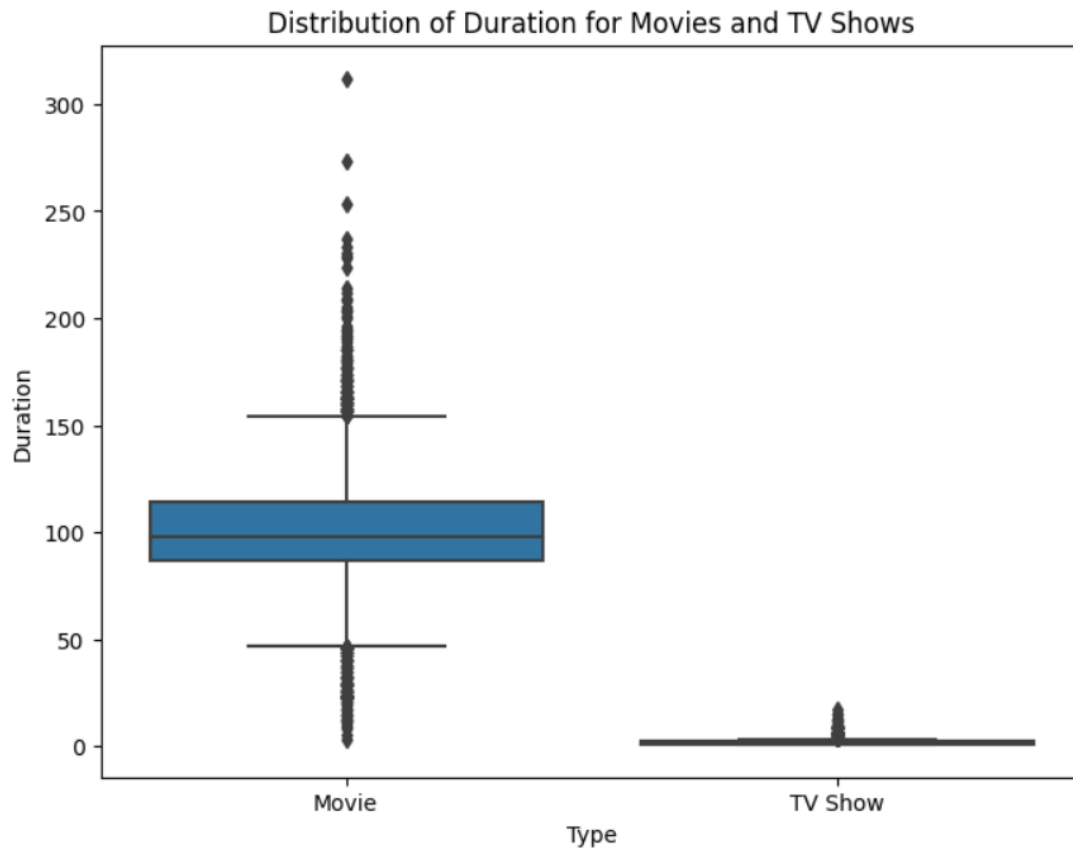
```
Country with less geners available : 77
Length of all the countries : 127
```

Observation :

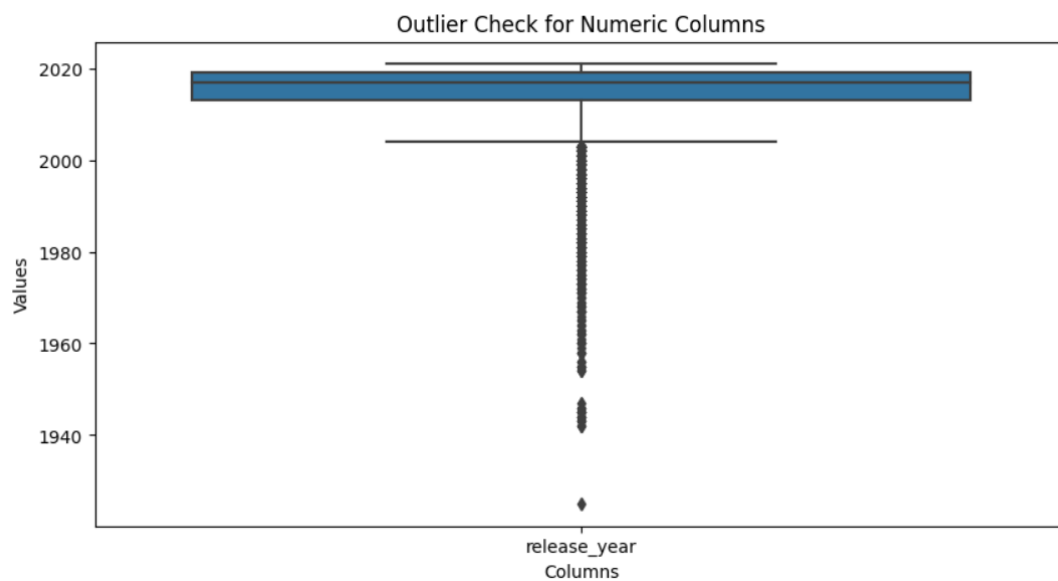
1. United States has highest number of unique content available followed by India, UK, France, and Canada.
2. International Movies is the most famous genre in the world as this is present in 2752 movies.
3. 77 countries out of 127 countries has less than 10 types of content movies available.

Outlier Check :

```
net['duration'] =  
net['duration'].str.extract('(\d+)').astype(float)  
plt.figure(figsize=(8, 6))  
sns.boxplot(x='type', y='duration', data=net)  
plt.title("Distribution of Duration for Movies and TV Shows")  
plt.xlabel("Type")  
plt.ylabel("Duration")  
plt.show()
```



```
numeric_columns = ['release_year', 'duration']
plt.figure(figsize=(10, 5))
sns.boxplot(data=netflix[numeric_columns])
plt.title("Outlier Check for Numeric Columns")
plt.xlabel("Columns")
plt.ylabel("Values")
plt.show()
```



Observation :

1. The box plot above shows the distribution of type over duration.
2. According to above plot we can say that the median of the movie data is ~100 where as almost all the movie duration is between 150 and 50.
3. There are some outliers below 50 and above 150 minutes. the maximum duration of the movie is more than 300 min.

Missing Value Check :

```
print(netflix.isnull().sum())
```

```
show_id      0
type         0
title        0
director    2634
cast        825
country     831
date_added   10
release_year  0
rating       4
duration     3
listed_in    0
description  0
dtype: int64
```

```
dire_missing_value = netflix.loc[netflix["director"].isna()]
A = dire_missing_value.groupby("type")["title"].count()
print(A)
```

```
Movie      188
TV Show    2446
```

Observation :

1. Data has missing values in director, cast, country, and date added column.

2. I have ignored the missing values in cast, country, and date added column as they are less in numbers for that particular columns.

3. For director Movie has very less missing values than TV Show. I am ignoring missing values in this column as well

since there are very less director's present in number of TV show directed, "Ken Burns" and "Alastair Fothergill"

are the only director's who have produced maximum TV Shows that is 3 ("Refer Q4"), So we can't replace the missing value

with them because here the missing value for TV Show is 2446.

Business Insights :

1. We have most missing values in "director" , "cast" and "country" column as per dataframe information.

2. Netflix has less old released movies than new released movies on platform.

3. Netflix has more movies that were released from 2015 to 2020 but there is sudden drop in the movie release in 2021 and

hence less movie addition on platform this is because of the pandemic. We can conclude from this is that people like

to watch more movies than a TV Show. One reason could be that Movies are having average duration of ~100 minutes and TV Shows are having ~2 seasons,

due to more time to complete a TV Show than Movies people prefer to watch Movies over Shows.

4. The best time to launch a TV Show is the 3rd quarter of the year which is July, August, September and the best month

is December as dec is the festival month in the world especially in the western region.

5. Maximum movies got added on Netflix in 2019 and after 2019 there is a less addition of movies in every year.

6. There is no sudden fall in addition of TV Shows on Netflix after 2019. TV Show is almost static after 2019.

7. After comparing both Movies and TV Show plot we can say after 2019 less movies got added on Netflix as compare to TV Show with respect to their previous years.

8. 77 countries out of 127 countries has less than 10 types of content movies available.

```
t6 = netflix.copy()
t6 = t6.dropna(subset = 'country')

def str_split(x):
    return str(x).split(', ')

def str_remove(x):
    x = x.replace('[', '')
    x = x.replace(']', '')
    return x

t6['country'] = t6['country'].apply(str_split)
t6 = t6.explode('country')
t6['country'] = t6['country'].apply(str_remove)
arr = t6['country'].value_counts().iloc[:10].index
t6 = t6[t6['country'].isin(arr)]
sns.countplot(x = 'country', data = t6, hue = 'type')
plt.xticks(rotation = 90)
plt.show()

t7 = t6[(t6['type'] == 'TV Show') & (t6['country'] ==
'India')]
sns.barplot(x = t7['release_year'].value_counts().index, y =
t7['release_year'].value_counts())
plt.show()

rat=nf_df['rating'].nunique(dropna=False)
rat
18

nf_df['rating'].unique()

['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', ..., '66 min',
'NR', NaN, 'TV-Y7-FV', 'UR']
Length: 18
```

```
Categories (17, object): ['66 min', '74 min', '84 min', 'G',
..., 'TV-Y', 'TV-Y7', 'TV-Y7-FV', 'UR']
nf_df['rating'].value_counts() >6
```

```
TV-MA True
TV-14 True
TV-PG True
R True
PG-13 True
TV-Y7 True
TV-Y True
PG True
TV-G True
NR True
G True
TV-Y7-FV False
UR False
NC-17 False
74 min False
84 min False
66 min False
```

```
Name: rating, dtype: bool
```

```
nf_df['release_year'].value_counts().head(10)
```

```
2018 1147
2017 1032
2019 1030
2020 953
2016 902
2021 592
2015 560
2014 352
2013 288
2012 237
```

```
Name: release_year, dtype: int64
```

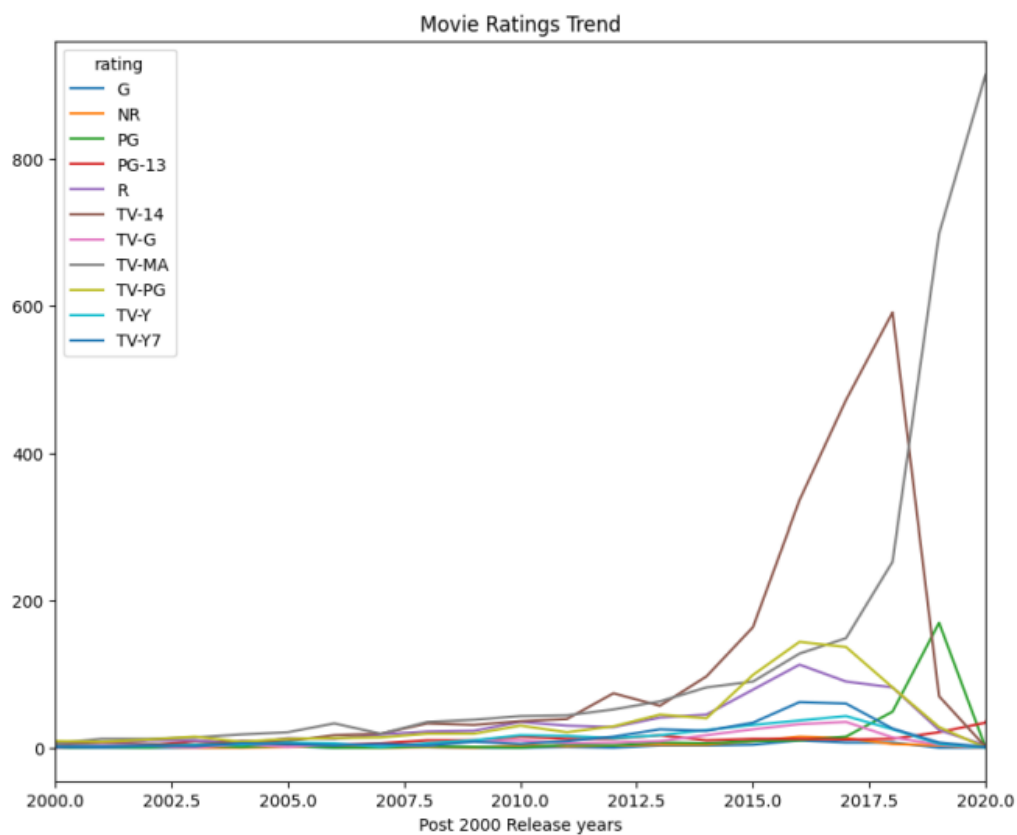
```
nf_rating_copy = nf_df
nf_rat = nf_df.groupby('rating')[['release_year']].count()
nf_rat = nf_rat.drop(['74 min', '84 min', '66 min', 'UR', 'TV-Y7-
FV', 'NC-17'])
nf_rat.reset_index(inplace = True)
nf_rat.columns = ['rating', 'release_year counts']
nf_rat
```

	rating	release_year	counts
0	G		41
1	NR		80
2	PG		287
3	PG-13		490
4	R		799
5	TV-14		2160
6	TV-G		220
7	TV-MA		2227

```

nf_rating=nf_rating_copy.merge(nf_rat, how='inner')
nf_rating.head()
nf_rating.shape
(8788, 15)
pd.crosstab(nf_df['release_year'],nf_rating['rating']
).plot(kind= 'line',figsize = (10,8),title = 'Movie Ratings
Trend')
plt.xlim(2000,2020)
plt.xlabel('Post 2000 Release years')
plt.show()

```



Recommendations :

1. The craze for TV Shows are increasing in India in recent years. Netflix can produce some good TV Show content to target

the massive Indian population.

2. Netflix should focus on increasing genres in movies in other countries other than top countries as out of 127 countries 77

countries have genre type less than 10 this will attract more customers to watch the content in countries where viewership is less.

3. As the values/counts were almost non changing for years less than 2000, I have visualized after 2000. From above line

plot, as we can clear see that TV-14(unsuitable for childer under 14) has been decreasing lately due to advancements in internet technology

and TV-MA(content for mature adults) has been more preferred now a days owing to lockdown restriction and self-isolations due to COVID. Hence

Netflix should focus more on content related to TV-MA ratings.

4. Out of total 962 Indian movies, most of the movies have duration in 100-150 mins. Netflix should focus on this range while

producing more in this range as evident from above PairPlot. Movies with rating TV-14(547) that is content for children above 14 are mostly

preferred as opposed to the overall movie ratings throughout the world where TV-MA(232) is dominant. Netflix should focus more on this

aspect with respect to Indian audience.

5. Out of total 84 tv shows, the tv shows which are mostly enjoyed and hence has more seasons as compared to others are

given in nf_India_tv dataframe. As it's evident from the PairPlot for TV Shows, Netflix should focus mostly on TV-MA (34) and TV-14 (25)

that is shows for mature adults and under 14 years as they are mostly preferred by the audience.