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

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**

1. Defining Problem Statement and Analysing basic metrics

Import Libraries

Importing the libraries we need

In [ ]:

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

Loading The Dataset

In [ ]:

```
#Using Pandas Library, we’ll load the CSV file. Named it with netflix_df for the dataset.
netflix_df = pd.read_csv("netflix_data.csv")
```

Let’s check the first 5 data.

In [ ]:

```
netflix_df.head()
```

Out[ ]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	des
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	n er
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	c pa Cap
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To hi p dr
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	fli ar
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In c k

In [ ]:

```
netflix_df
```

Out[ ]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV

Orleans

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...
...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals

8807 rows × 12 columns

The dataset contains over 8807 rows, 12 columns. After a quick view of the data frames, it looks like a typical movie/TVshows data frame without ratings. We can also see that there are NaN values in some columns.

2: Observations on the shape of data, data types of allthe attributes, conversion of categorical attributes to'category' (If required), missing value detection,statistical summary

In [ ]:

```
# To get All attributes netflix_df.columns
netflix_df.columns
```

Out[ ]:

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

In [ ]:

```
netflix_df.shape
```

Out[ ]:

Out[ ]:

```
(8807, 12)
```

In [ ]:

```
# Data types of all the attributes
netflix_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8807 non-null   object
 1   type          8807 non-null   object
 2   title         8807 non-null   object
 3   director      6173 non-null   object
 4   cast          7982 non-null   object
 5   country       7976 non-null   object
 6   date_added    8797 non-null   object
 7   release_year  8807 non-null   int64 
 8   rating        8803 non-null   object
 9   duration      8804 non-null   object
 10  listed_in     8807 non-null   object
 11  description   8807 non-null   object
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
```

In [ ]:

```
# Statistical Summary Before Data Cleaning:
netflix_df.describe()
```

Out[ ]:

	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

Missing Value Detection

Data Profiling & Cleaning

Data Cleaning means the process of identifying incorrect, incomplete, inaccurate, irrelevant, or missing pieces of data and then modifying, replacing, or deleting them as needed. Data Cleansing is considered as the basic element of Data Science.

In [ ]:

```
print('\nColumns with missing value:')
print(netflix_df.isnull().any())
```

```
Columns with missing value:
show_id         False
type            False
title           False
director         True
cast             True
```

```
country           True
date_added        True
release_year     False
rating            True
duration          True
listed_in        False
description      False
dtype: bool
```

From the info, we know that there are 8807 entries and 12 columns to work with for this EDA. There are a few columns that contain null values, “director,” “cast,” “country,” “date_added,” “rating.”

In [ ]:

```
netflix_df.T.apply(lambda x: x.isnull().sum(), axis = 1)
```

Out[ ]:

	0
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

In [ ]:

```
netflix_df.isnull().sum().sum()
```

Out[ ]:

```
4307
```

There are a total of 4307 null values across the entire dataset with 2634 missing points under "director", 825 under "cast", 831 under "country", 11 under "date_added", 4 under "rating" and 3 under “duration ”. We will have to handle all null data points before

---

we can dive into EDA and modelling.

Imputation is a treatment method for missing value by filling it in using certain techniques.

Can use mean, mode, or use predictive modelling. In thiscase study, we will discuss the use of the fillna function from Pandas for this imputation. Drop rows containing missing values. Can use the dropna function from Pandas.

In [ ]:

```
netflix_df.director.fillna("No Director", inplace=True)
netflix_df.cast.fillna("No Cast", inplace=True)
netflix_df.country.fillna("Country Unavailable", inplace=True)
netflix_df.dropna(subset=["duration", "rating"], inplace=True)
```

Checking Missing values

In [ ]:

```
netflix_df.isna().any()
```

Out[ ]:

	0
show_id	False
type	False
title	False
director	False
cast	False
country	False
date_added	True
release_year	False
rating	False
duration	False
listed_in	False
description	False

dtype: bool

For missing values, the easiest way to get rid of them would be to delete the rows with the missing data. However, this wouldn't be beneficial to our EDA since the is a loss of information. Since "director", "cast", and "country" contain the majority of null values, we chose to treat eachmissing value is unavailable. The other two label "date_added”,” duration” and "rating" contain an insignificant portion of the data so it drops from the dataset. Finally, we can see that there are no more missing values in the data frame.

Statistical Summary After Data Cleaning:

In [ ]:

```
netflix_df.describe()
```

Out[ ]:

	release_year
count	8800.000000
mean	2014.179886
std	8.822583
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

3. Non-Graphical Analysis

Non-Graphical Analysis involves calculating the summary statistics, without using pictorial or graphical representations. There are 3 main functions that Pandas library provide us, and I will be discussing about them. Those functions are:

1. **info()**
2. **isna().sum() or isnull().sum()**
3. **describe()**

In [ ]:

```
# Checking the data using .head()
netflix_df.head()
```

Out[ ]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	No Cast	United States	September 25, 2021	2020	PG-13	90 min	Documentaries
1	s2	TV Show	Blood & Water	No Director	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Country Unavailable	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
3	s4	TV Show	Jailbirds New Orleans	No Director	No Cast	Country Unavailable	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV
4	s5	TV Show	Kota Factory	No Director	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...

1.info() mainly indicates the number of features, non-null count, and data type of each features. Additionally, it also shows the number of features in present in each data type(s). This helps us to determine how many numerical and categorical features we have.

In [ ]:

```
netflix_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 8800 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   show_id        8800 non-null   object
 1   type           8800 non-null   object
 2   title          8800 non-null   object
 3   director      8800 non-null   object
 4   cast          8800 non-null   object
 5   country       8800 non-null   object
 6   date_added    8790 non-null   object
 7   release_year  8800 non-null   int64
 8   rating        8800 non-null   object
 9   duration      8800 non-null   object
 10  listed_in     8800 non-null   object
 11  description   8800 non-null   object
dtypes: int64(1), object(11)
memory usage: 893.8+ KB
```

2.Read The Description Of The Data

In [ ]:

```
netflix_df.describe()
```

Out[ ]:

	release_year
count	8800.000000
mean	2014.179886
std	8.822583
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

1. isna().sum() or isnull().sum()

In [ ]:

```
netflix_df.T.apply(lambda x: x.isnull().sum(), axis = 1)
```

Out[ ]:

	0
show_id	0
type	0
title	0
director	0
cast	0
country	0
date_added	10
release_year	0
rating	0
duration	0
listed_in	0
description	0

dtype: int64

In [ ]:

```
netflix_movies_df = netflix_df[netflix_df.type == "Movie"]
netflix_shows_df = netflix_df[netflix_df.type == "TV Show"]
netflix_movies_df.head()
```

Out[ ]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	liste
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	No Cast	United States	September 25, 2021	2020	PG-13	90 min	Document
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	Country Unavailable	September 24, 2021	2021	PG	91 min	Childr Family M
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	Dra Indeper Mo Internat M
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG-13	104 min	Come Dra
12	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	2021	TV-MA	127 min	Dra Internat M

In [ ]:

```
netflix_shows_df.head()
```

Out[ ]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	de
1	s2	TV Show	Blood & Water	No Director	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	Ca
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Country Unavailable	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	T
3	s4	TV Show	Jailbirds New Orleans	No Director	No Cast	Country Unavailable	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	de
4	s5	TV Show	Kota Factory	No Director	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	I
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	Country Unavailable	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	T ch

4: Exploratory Analysis and Visualization

Visual Analysis - Univariate, Bivariate after preprocessing of the data

Univariate analysis

Analysis done based only on one variable. we are not going to the math behind these concepts, for now, let's see what these are in graphs.

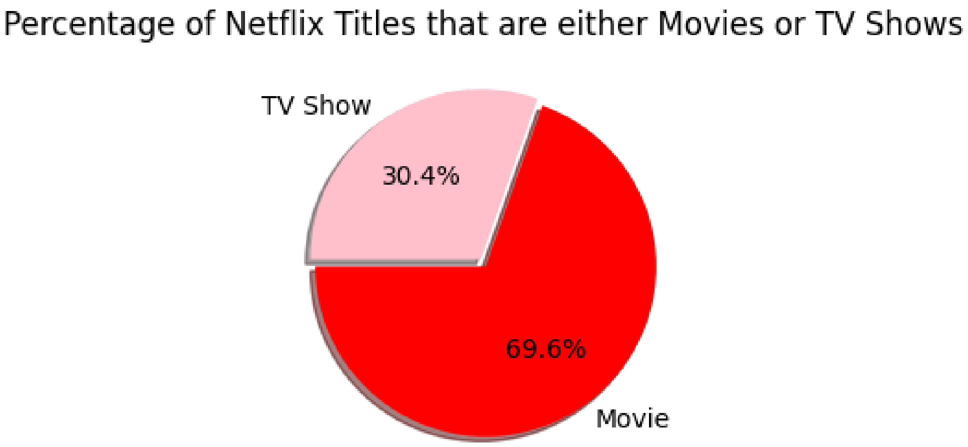
A==>Pie plot:

Netflix Content By Type

Analysis entire Netflix dataset consisting of both movies and shows. Let’s compare the total number of movies and shows in this dataset to know which one is the majority

In [ ]:

```
plt.figure(figsize=(6,3))
plt.title("Percentage of Netflix Titles that are either Movies or TV Shows")
g=plt.pie(netflix_df.type.value_counts(),explode=(0.025,0.025),
labels=netflix_df.type.value_counts().index, colors=['red','pink'],autopct='%1.1f%%',
startangle=180,shadow=True)
plt.show()
```



There are far more movie titles (69.7%) that TV shows titles (30.3%) in terms of title.

2. Amount of Content as a Function of Time:

Lineplot we will explore the amount of content Netflix has added throughout the previous years. Since we are interested in when Netflix added the title onto their platform, we will add a “year_added” column to show the date from the “date_added” columns.

In [ ]:

```
# Handle potential inconsistencies in date formatting
netflix_df["year_added"] = pd.to_datetime(netflix_df.date_added.str.strip(), format='%B
%d, %Y', errors='coerce').dt.year
netflix_movies_df["year_added"] = pd.to_datetime(netflix_movies_df.date_added.str.strip(
), format='%B %d, %Y', errors='coerce').dt.year
netflix_shows_df["year_added"] = pd.to_datetime(netflix_shows_df.date_added.str.strip(),
format='%B %d, %Y', errors='coerce').dt.year
netflix_year_df = netflix_df.year_added.value_counts().to_frame().reset_index().rename(c
olumns={"index": "count",
"year_added":"year"})
netflix_year_df = netflix_year_df[netflix_year_df['year'] != 2020]
print(netflix_year_df)
```

```
      year  count
0   2019.0   2016
2   2018.0   1648
3   2021.0   1498
4   2017.0   1185
5   2016.0    426
6   2015.0     82
7   2014.0     24
8   2011.0     13
9   2013.0     11
10  2012.0      3
11  2009.0      2
12  2008.0      2
13  2010.0      1
```

```
<ipython-input-22-6e0a5f9cc98c>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_g
uide/indexing.html#returning-a-view-versus-a-copy
  netflix_movies_df["year_added"] = pd.to_datetime(netflix_movies_df.date_added.str.strip
(), format='%B %d, %Y', errors='coerce').dt.year
<ipython-input-22-6e0a5f9cc98c>:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_g
uide/indexing.html#returning-a-view-versus-a-copy
  netflix_shows_df["year_added"] = pd.to_datetime(netflix_shows_df.date_added.str.strip()
, format='%B %d, %Y', errors='coerce').dt.year
```

In [ ]:

```
movies_year_df =netflix_movies_df.year_added.value_counts().to_frame().reset_index().ren
ame(columns={"index":
"count", "year_added":"year"})
movies_year_df = movies_year_df[movies_year_df['year'] != 2020]
movies_year_df
```

Out[ ]:

	year	count
0	2019	1424
2	2018	1237
3	2021	993
4	2017	836
5	2016	251
6	2015	56
7	2014	19
8	2011	13

	year	count
9	2013	6
10	2012	3
11	2009	2
12	2008	1
13	2010	1

In [ ]:

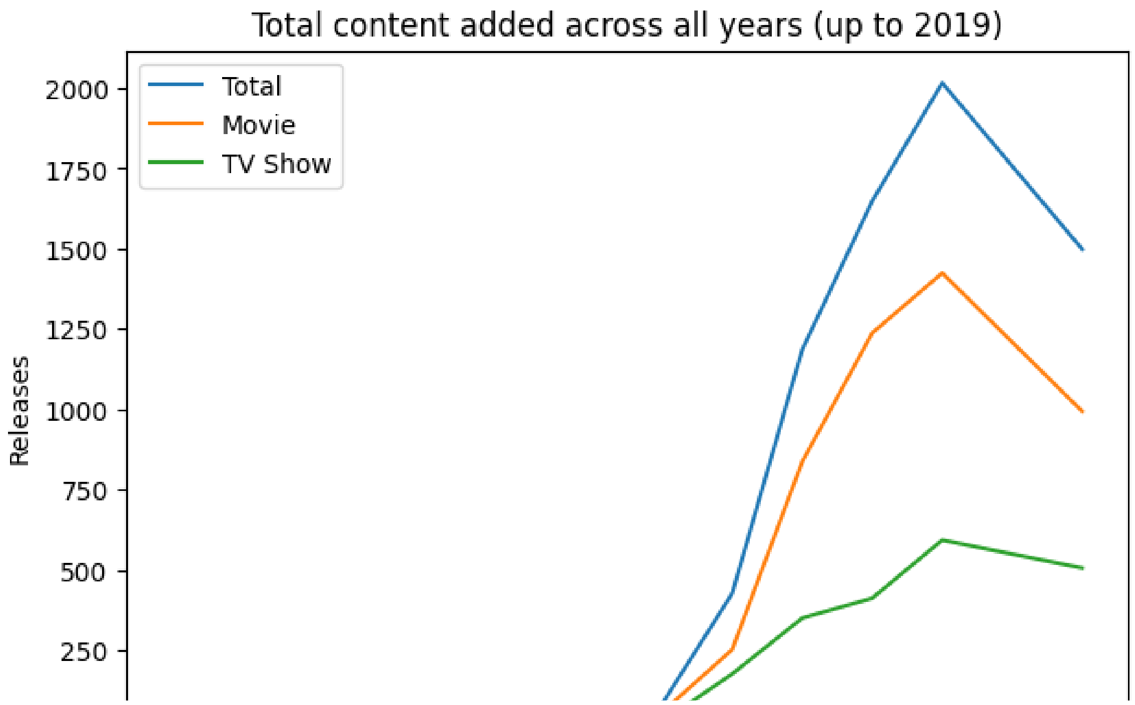
```
shows_year_df = netflix_shows_df.year_added.value_counts().to_frame().reset_index().rena
me(columns={"index":
"count", "year_added":"year"})
shows_year_df = shows_year_df[shows_year_df['year'] != 2020]
shows_year_df
```

Out[ ]:

	year	count
1	2019.0	592
2	2021.0	505
3	2018.0	411
4	2017.0	349
5	2016.0	175
6	2015.0	26
7	2014.0	5
8	2013.0	5
9	2008.0	1

In [ ]:

```
fig, ax = plt.subplots(figsize=(7, 5))
sns.lineplot(data=netflix_year_df, x='year', y='count', label='Total')
sns.lineplot(data=movies_year_df, x='year', y='count', label='Movie')
sns.lineplot(data=shows_year_df, x='year', y='count', label='TV Show')
ax.set_xticks(np.arange(2008, 2020, 1))
plt.title("Total content added across all years (up to 2019)")
plt.ylabel("Releases")
plt.xlabel("Year")
plt.show()
```



Based on the timeline above, we can conclude that the popular streaming platform started gaining traction after 2013. Since then, the amount of content added has been increasing significantly. The growth in the number of movies on Netflix is much higher than that on TV shows. About 1,300 new movies were added in both 2018 and 2019. Besides, we can know that Netflix has increasingly focused on movies rather than TV shows in recent years

3. Exploring the countries contribution with the most content of Netflix.

Next is exploring the countries by the amount of the produces content of Netflix. We need to separate all countries within a film before analysing it, then removing titles with no countries available.

In [ ]:

```
import plotly.graph_objects as go
from plotly.offline import init_notebook_mode, iplot
```

We need to separate all countries within a film before analyzing it, then removing titles with no countries available

In [ ]:

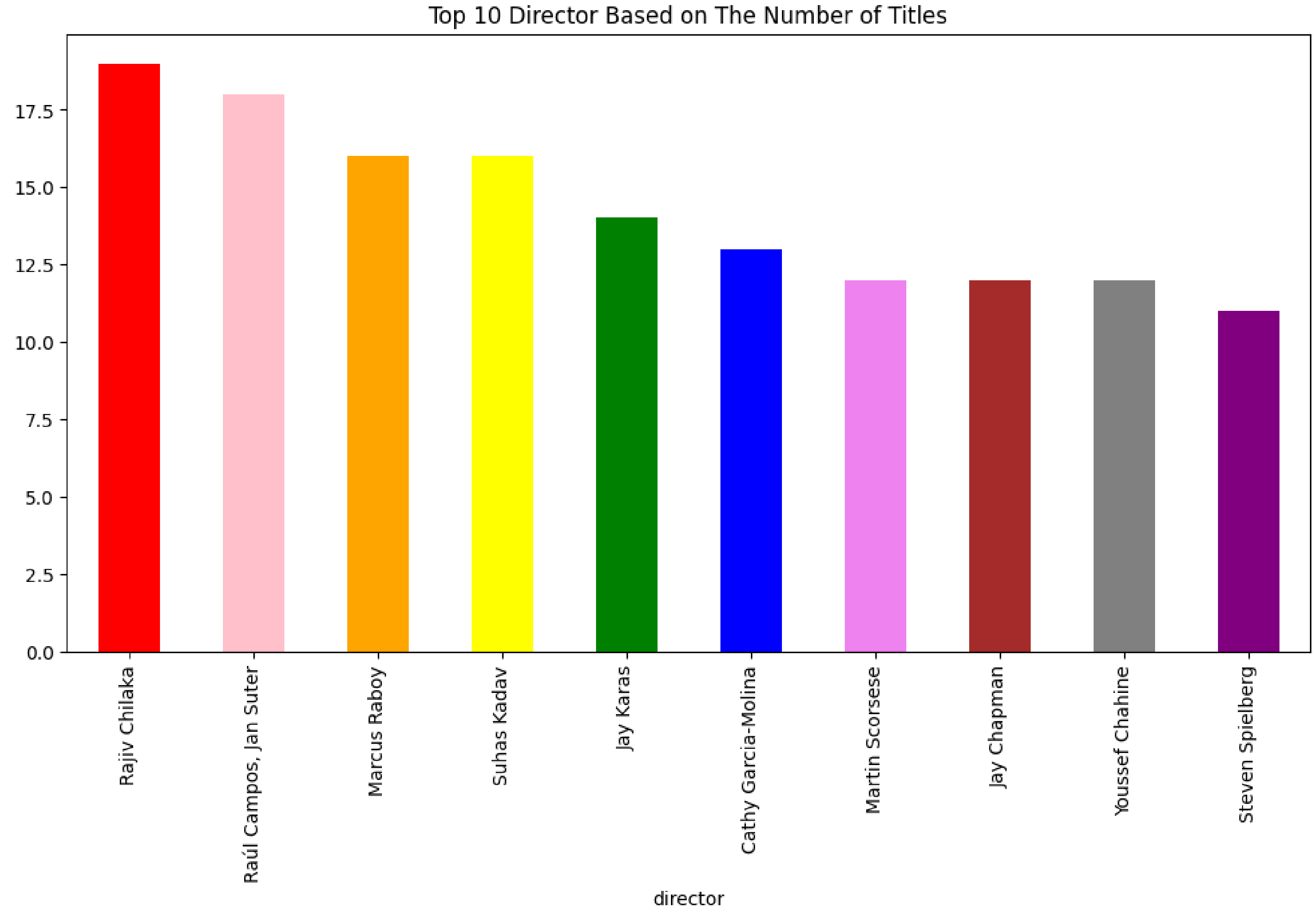
```
filtered_countries = netflix_df.set_index('title').country.str.split(', ', expand=True).
stack().reset_index(level=1, drop=True);
filtered_countries = filtered_countries[filtered_countries != 'Country Unavailable']
iplot([go.Choropleth(locationmode='country names',locations=filtered_countries,z=filtered
_countries.value_counts())])
```

4. Top Directors on Netflix

To know the most popular director, we can visualize it.

In [ ]:

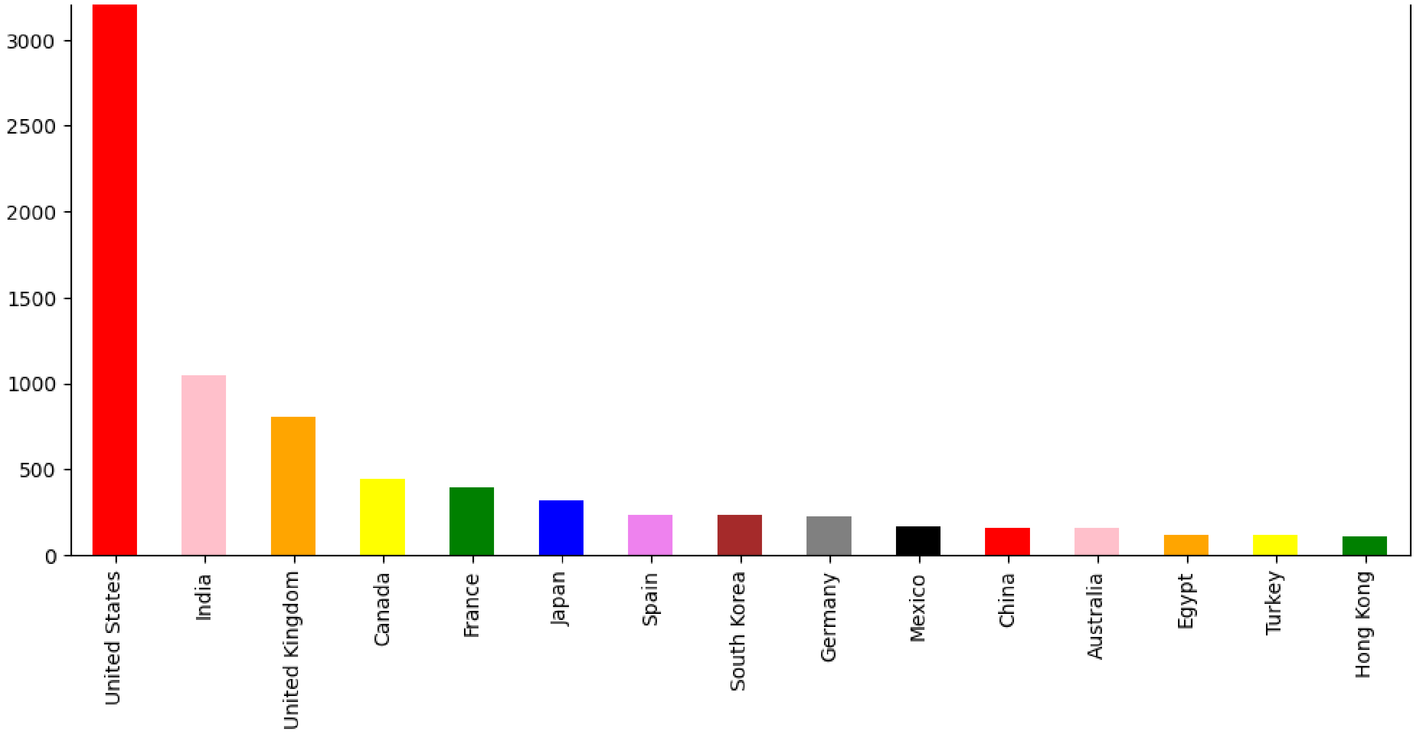
```
plt.figure(figsize=(12,6))
netflix_df[netflix_df["director"] != "No Director"]["director"].value_counts().head(10).plot(kind="bar",color=['red','pink','orange','yellow','green','blue','violet','brown','gray','purple'])
plt.title("Top 10 Director Based on The Number of Titles")
plt.show()
```



- **Raul Campos and Jan Suter are the most prolific directors on Netflix, having directed the highest number of titles.**
- **This indicates that Netflix might have a preference for certain directors or that these directors have a successful track record on the platform.**
- **It is important to note that this graph only shows the number of titles directed by each person, not their popularity or viewership.**
- **Analyzing the genres and types of content these directors produce could provide further insights into Netflix's content strategy.**

In [ ]:

```
plt.figure(figsize=(12,6))
filtered_countries.value_counts().head(15).plot(kind="bar",color=['red','pink','orange','yellow','green','blue','violet','brown','gray','black'])
plt.title("Top 10 Countries with most content")
plt.show()
```



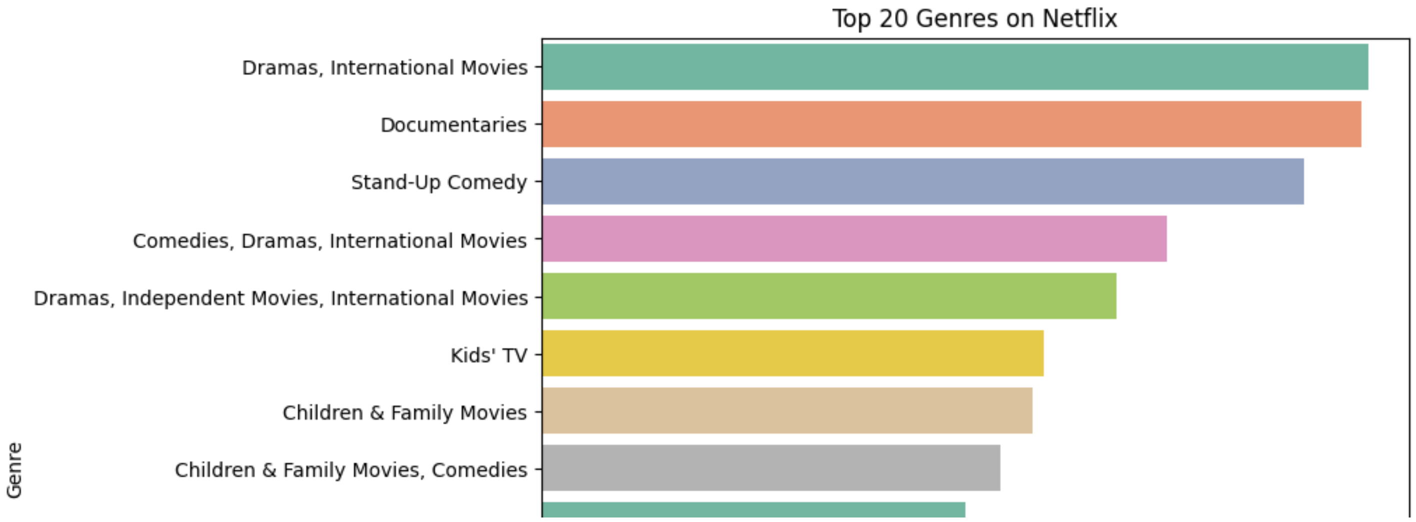
- **The United States has the highest number of titles on Netflix, followed by India and the United Kingdom.**
- **This suggests that Netflix has a strong focus on content produced in these countries.**
- **However, it's important to note that this data may not reflect the popularity of content in different countries.**
- **For example, a title produced in the United States may be more popular in India than a title produced in India.**

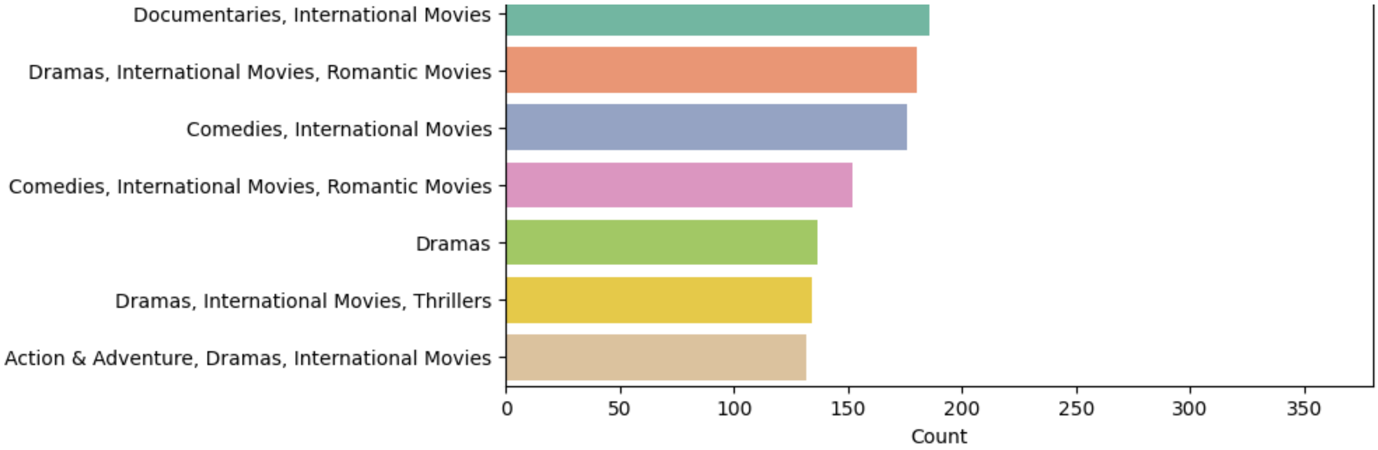
5. Top 20 Genres on Netflix: Count Plot

In [ ]:

```
plt.figure(figsize=(8,8))
ax = sns.countplot(y="listed_in", data=netflix_df, order=netflix_df.listed_in.value_coun
ts().index[0:15], palette="Set2")
plt.title("Top 20 Genres on Netflix")
plt.xlabel("Count")
plt.ylabel("Genre")
plt.show()
```

```
<ipython-input-30-42a59155adf4>:2: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. A
ssign the `y` variable to `hue` and set `legend=False` for the same effect.
```



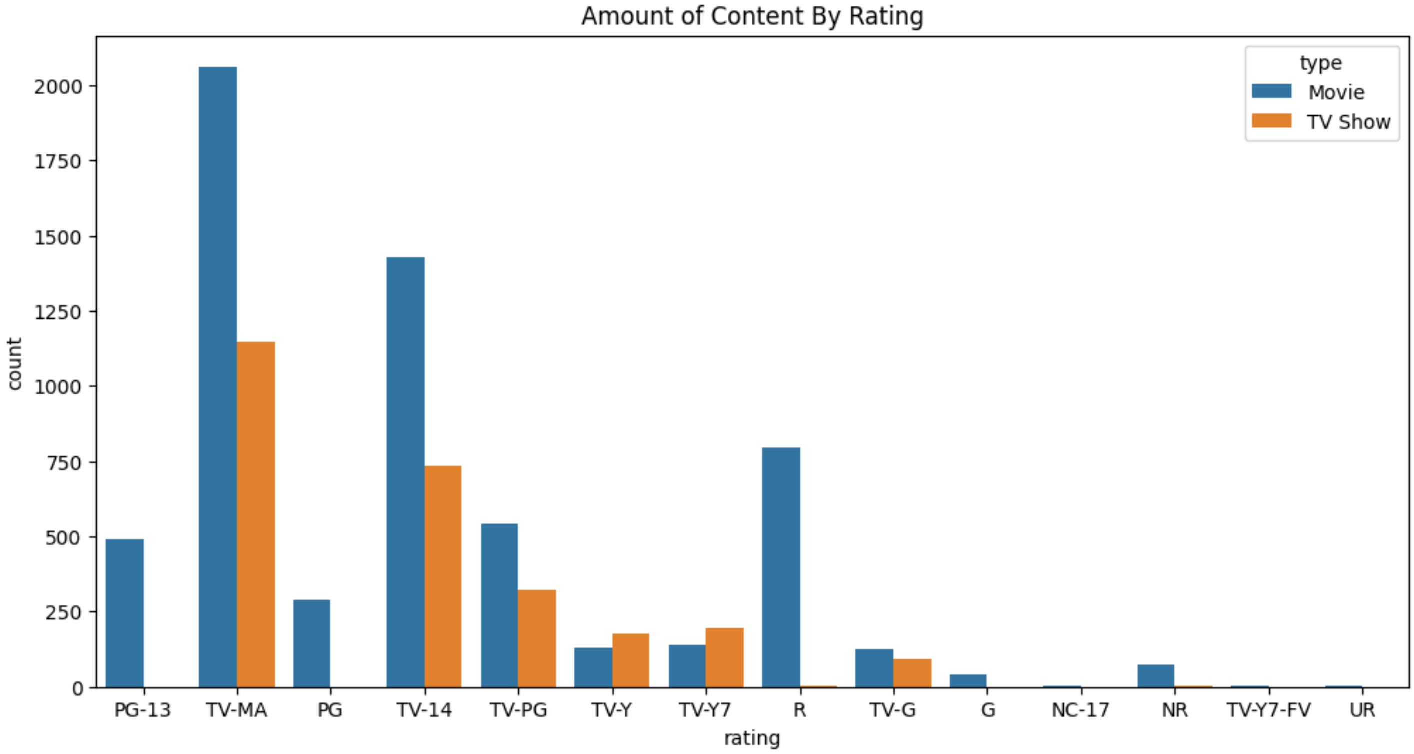


- The graph shows the top 20 genres on Netflix.
- We can see that the most popular genres are International Movies, Dramas, Comedies, and Documentaries.
- This suggests that Netflix has a wide variety of content to offer its users, and that it is focusing on providing content that appeals to a broad audience.
- This information can be used to inform content acquisition and production decisions.

Amount of Content By Rating

In [ ]:

```
plt.figure(figsize=(12,6))
sns.countplot(x='rating',hue='type',data=netflix_df)
plt.title("Amount of Content By Rating")
plt.show()
```



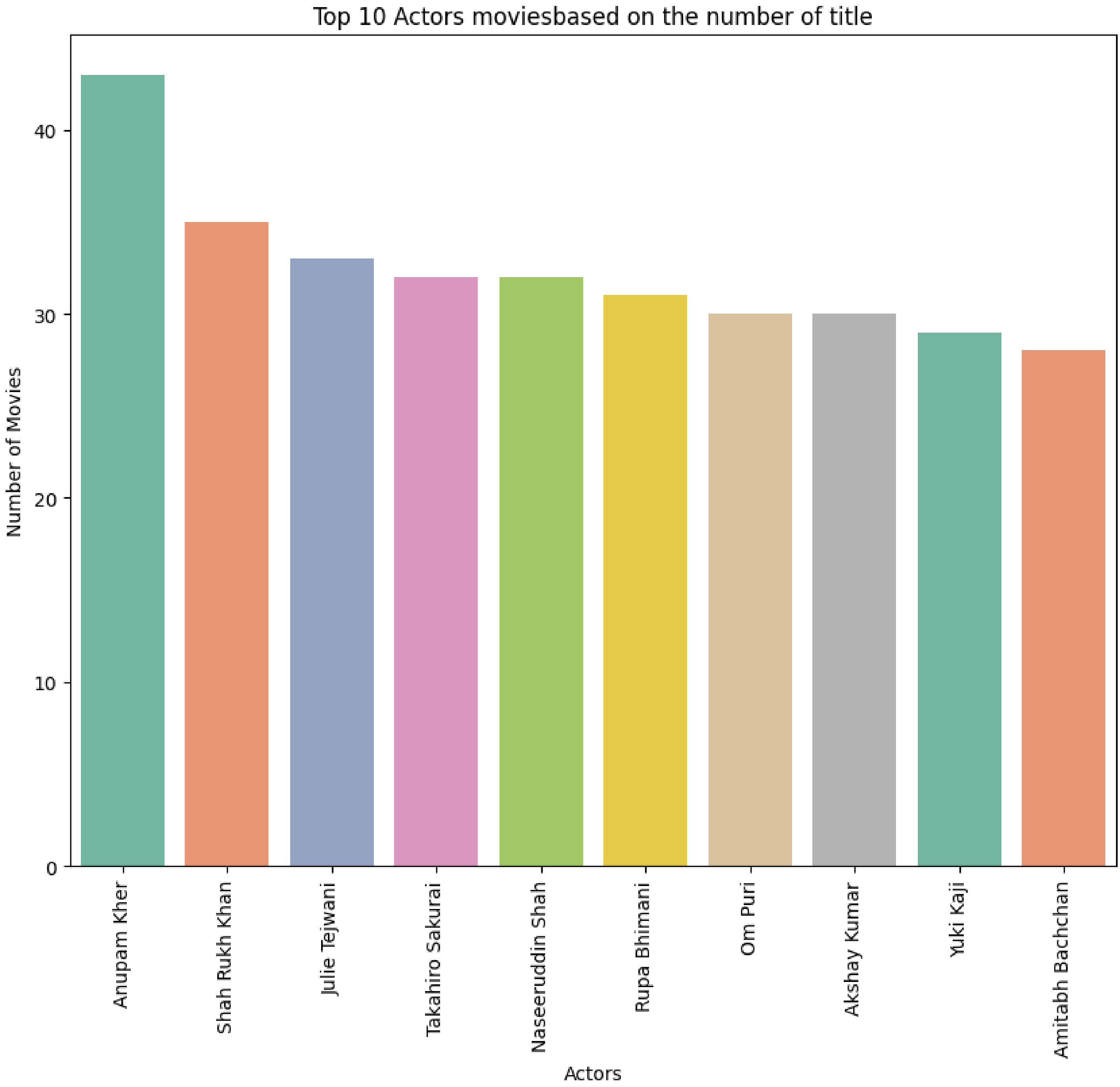
The majority of content on Netflix is categorized with a TV-MA rating, which indicates that it is intended for mature audiences only.This is followed by TV-14 content, which is suitable for viewers aged 14 and above. There is a significant amount of content available for younger audiences as well, with TV-Y7, TV-Y, and TV-G ratings representing a substantial portion of the offerings. The distribution of ratings suggests that Netflix caters to a wide range of viewers, with a particular focus on adult content. The platform also offers a diverse selection of movies and TV shows across various genres to appeal to different tastes and preferences.

In [ ]:

```
cast_shows = netflix_df[netflix_df["cast"] != "No Cast"].set_index("title").cast.str.spl
```

```
it(", ", expand=True).stack().reset_index(level=1, drop=True)
top_10_actors = cast_shows.value_counts().head(10)

plt.figure(figsize=(10, 8))
sns.barplot(x=top_10_actors.index, y=top_10_actors.values,  palette='Set2')
plt.xticks(rotation=90)
plt.xlabel("Actors")
plt.ylabel("Number of Movies")
plt.title("Top 10 Actors moviesbased on the number of title")
plt.show()
```

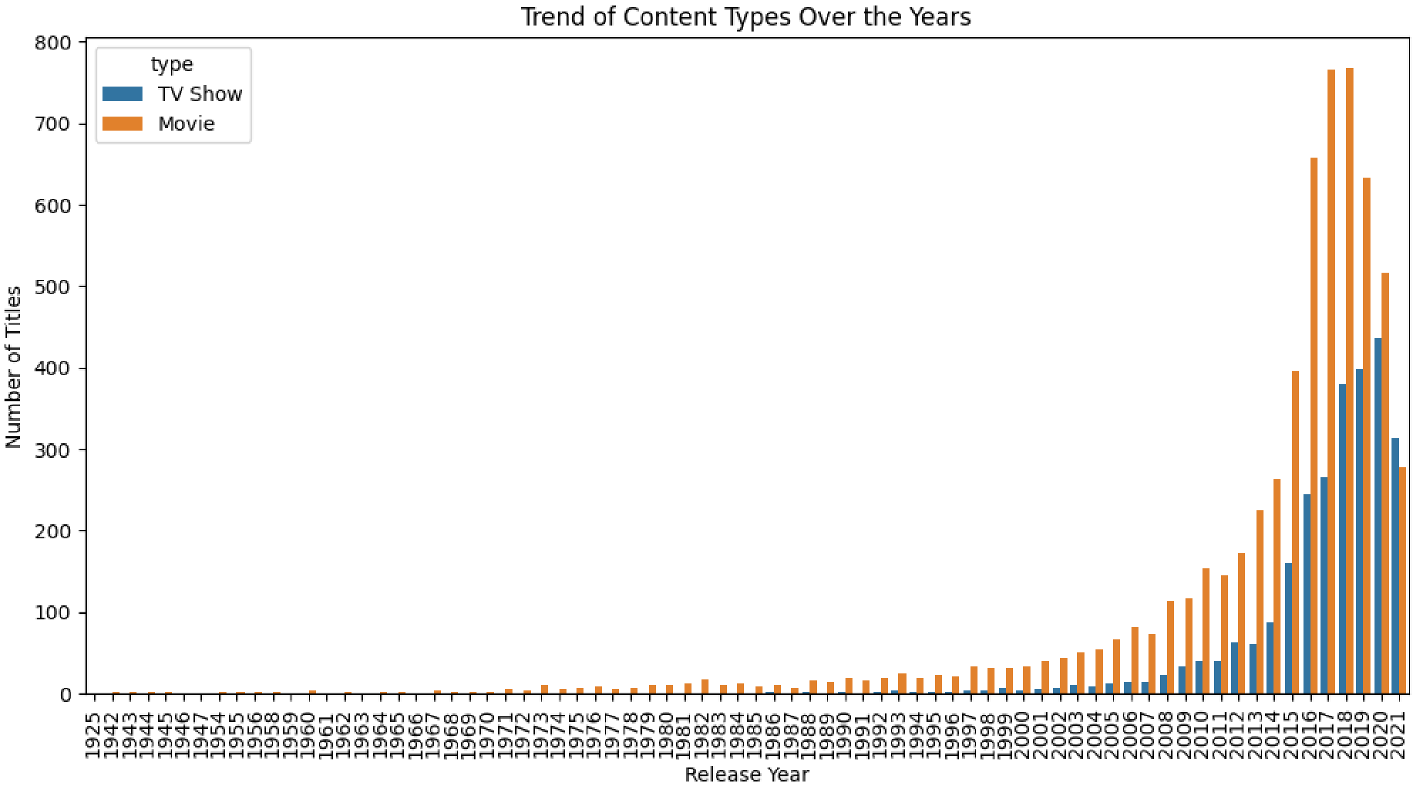
```
<ipython-input-32-fe10bbbac824>:5: FutureWarning:



Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. A
ssign the `x` variable to `hue` and set `legend=False` for the same effect.
```



The top actor on Netflix Movies, based on the number of titles, is Anupam Kher, followed by Shahrukh Khan and Julie Tejwani

```
In [ ]:
```

```
plt.figure(figsize=(12,6))
sns.countplot(x='release_year', hue='type', data=netflix_df)
plt.xticks(rotation=90)
plt.title('Trend of Content Types Over the Years')
plt.xlabel('Release Year')
plt.ylabel('Number of Titles')
plt.show()
```



- **The graph shows the trend of content types (movies and TV shows) over the years.**
- **We can see that the number of movies released has been increasing steadily over the years, while the number of TV shows has been more volatile.**
- **This suggests that Netflix is increasingly focusing on movies, which could be due to the growing popularity of movies and the increasing availability of streaming services.**
- **It is also possible that Netflix is investing more in movie production than in TV shows.**

Conclusions:

Content Distribution: Netflix offers a significantly larger library of movies compared to TV shows.

Growth Trend: Netflix experienced substantial growth in content additions after 2013, with a particular focus on movies in recent years.

Geographic Focus: The United States is the primary contributor of content on Netflix, followed by India and the United Kingdom.

Popular Genres: International Movies, Dramas, and Comedies are among the most prevalent genres on the platform.

Content Rating: A majority of Netflix content is rated TV-MA, indicating a focus on mature audiences. However, the platform also offers a considerable selection for younger viewers.

Top Directors: Rajiv Chilaka, Raúl Campos, and Jan Suter are among the most prolific directors on Netflix.

Top Actors: Anupam Kher, Shahrukh Khan, and Julie Tejwani are the most frequently appearing actors in Netflix movies.

Release Year Trend: There's a clear trend of increasing content releases over the years, with a notable surge in movie releases in recent years.

Recommendations:

Content Diversification: While Netflix has a strong focus on movies, consider expanding the TV show library to cater to a broader audience. This could involve investing in original TV show productions across diverse genres.

Global Expansion: Leverage the popularity of international content by further investing in productions from countries like India, the United Kingdom, and other emerging markets. This could involve partnering with local production houses and talent.

Genre Focus: Continue to strengthen the platform's offerings in popular genres like International Movies, Dramas, and Comedies. Additionally, explore opportunities in niche genres to attract specific audience segments.

Content Rating Balance: While catering to mature audiences is important, consider expanding *content* for younger viewers and families. This could involve acquiring or producing more titles with TV-Y, TV-Y7, and TV-G ratings.

Talent Acquisition: Collaborate with top directors like Rajiv Chilaka, Raúl Campos, and Jan Suter, and leverage the star power of actors like Anupam Kher, Shahrukh Khan, and Julie Tejwani to attract viewership.

Data-Driven Decision Making: Utilize data analytics to identify emerging trends, understand audience preferences, and make informed decisions regarding content acquisition and production. This could involve analyzing viewing patterns, ratings, and social media sentiment.

Personalized Recommendations: Enhance the recommendation system to provide more personalized suggestions to users based on their viewing history and preferences. This could improve user engagement and retention.

Technological Innovation: Continue to invest in technological advancements to improve streaming quality, user interface, and overall viewing experience. This could involve exploring new formats like interactive content or virtual reality experiences.

In [ ]: