

# Netflix - Data Exploration and Visualisation

## Introduction:

### 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.

### Objective:

Analyzing 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.

### Features of the Dataset:

<b>Show_id:</b>	Unique ID for every Movie / Tv Show
<b>Type:</b>	Identifier - A Movie or TV Show
<b>Title:</b>	Title of the Movie / Tv Show
<b>Director:</b>	Director of the Movie
<b>Cast:</b>	Actors involved in the movie/show
<b>Country:</b>	Country where the movie/show was produced
<b>Date_added:</b>	Date it was added on Netflix
<b>Release_year:</b>	Actual Release year of the movie/show
<b>Rating:</b>	TV Rating of the movie/show
<b>Duration:</b>	Total Duration - in minutes or number of seasons
<b>Listed_in:</b>	Genre
<b>Description:</b>	The summary description

## Importing Libraries:

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

df = pd.read_csv("/content/netflix.csv")
df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descrip
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, film
1	s2	TV Show	Blood & Water	NaN	Ama Camata, Khosi Ngema, Gail Mababalane, Thabani...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party in Cape Town
2	s3	TV Show	Ganglands	Julien Leclercq	Sam Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk in a prison
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 a city of coaching centers known for its to tra

it steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

## Finding the DataFrame Dimension

```
: df.shape
: (8807, 12)
```

## Exploration of Data

```
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
```

**Insights:** By analysing the above information we conclude that data has 12 features in the combination of mixed alpha numeric data and there are some missing values as well which represents as NaN.

```
df.describe(include="object")
```

	show_id	type	title	director	cast	country	date_added	rating	duration	listed_in	description
count	8807	8807	8807	6173	7982	7976	8797	8803	8804	8807	8807
unique	8807	2	8807	4528	7692	748	1767	17	220	514	8775
top	s1	Movie	Dick Johnson Is Dead	Rajiv Chilaka	David Attenborough	United States	January 1, 2020	TV-MA	1 Season	Dramas, International Movies	Paranormal activity at a lush, abandoned prope...
freq	1	6131	1	19	19	2818	109	3207	1793	362	4

### Type column:

From the above data there are 8807 items available on Netflix. By observing the frequency in the above data which shows there are 6131 them are movies.

### Director column:

By observing the above data Rajiv Chilaka directed the most number of TV-Shows and Movies.

### Country column:

The USA country holds the first position where highest number TV-Show and Movies were added on the Netflix Platform.

### Data Added column:

This column data shows January 1,2020 is the most repeated in the data we can confirm that highest number of TV-Shows and movies were added on this date.

### Rating column:

There are 17 different types of ratings present on Netflix. The "TV-MA" (Mature Audience Only) rating dominates the charts, covering almost 36% of the total shows and movies on the platform with this rating. We can conclude that shows and movies added the Netflix and adult content.

## Adding Extra columns for better analysis

```
[6] df["date_added"] = df["date_added"].str.strip()

[7] import pandas as pd
    df["year_added"] = pd.to_datetime(df["date_added"]).dt.year

[8] df["month_added"] = pd.to_datetime(df["date_added"]).dt.month

[9] df["day_added"] = pd.to_datetime(df["date_added"]).dt.day_name()

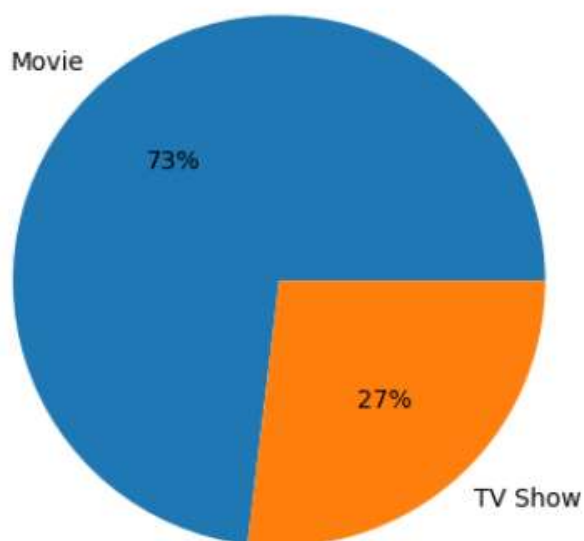
[10] df["month_name"] = pd.to_datetime(df["year_added"]).dt.month_name()
```

**Insights:** For better analysis extracting year, month, day\_name, month\_name from date\_added column. Now we can do some in depth analysis on the data by taking the following information into consideration

```
dt_ty = df["type"].value_counts()
dt_ty

import matplotlib.pyplot as plt
import seaborn as sns

plt.pie(dt_ty,
        labels = dt_ty.index,
        autopct = "%.f%%"
    )
plt.show()
```



**Insights:** From the above information and pie chart we can conclude that there are 6131 movies as 73% in the whole data and 2676 TV Shows as 27% . Movies are dominating on the Netflix platform when comparing to the TV Show which are emerging newly.

### Checking any Duplicates present in the data:

	<code>df.duplicated().value_counts()</code>
	count
False	8807
dtype: int64	

**Insights:** By observing the above information we can conclude that there are no duplicate found in the dataset.

```
df_ty2=df.isnull().sum()
df_ty2
```

	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

## Finding the missing values in form of percentages

```
[59] for i in df.columns:
      null_rate = df[i].isnull().sum()/df.shape[0] * 100
      if null_rate > 0:
          print(f"{i}'s null rate : {round(null_rate,2)}%")
```

```
director's null rate : 29.91%
cast's null rate : 9.37%
country's null rate : 9.44%
date_added's null rate : 0.11%
rating's null rate : 0.05%
duration's null rate : 0.03%
```

### Insights:

By observing the above data directors column as most null values followed by cast and country others contain very less.

## Un-nesting the columns for better analysis:

```

) df["director"] = df["director"].str.split(", ")
df["cast"] = df["cast"].str.split(", ")
df["country"] = df["country"].str.split(", ")

df = df.explode(['director'])
df = df.explode(['cast'])
df = df.explode(['country'])
df

```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons
1	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	2 Seasons
1	s2	TV Show	Blood & Water	NaN	Gail Mabalane	South Africa	September 24, 2021	2021	TV-MA	2 Seasons
1	s2	TV Show	Blood & Water	NaN	Thabang Molaba	South Africa	September 24, 2021	2021	TV-MA	2 Seasons
...	...	...	...	...	...	...	...	...	...	...
8806	s8807	Movie	Zubaan	Mozes Singh	Manish Chaudhary	India	March 2, 2019	2015	TV-14	111 min
8806	s8807	Movie	Zubaan	Mozes Singh	Meghna Malik	India	March 2, 2019	2015	TV-14	111 min
8806	s8807	Movie	Zubaan	Mozes Singh	Malkeet Rauni	India	March 2, 2019	2015	TV-14	111 min
8806	s8807	Movie	Zubaan	Mozes Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 min
8806	s8807	Movie	Zubaan	Mozes Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 min

89382 rows × 12 columns

Filling the missing values with unknown values :

```

df["director"].fillna("Unknown_director",inplace=True)
df["cast"].fillna("Unknown_cast",inplace=True)
df["country"].fillna("Unknown_country",inplace=True)
df["date_added"].fillna("Unknown_date",inplace=True)
df["rating"].fillna("Unknown_rating",inplace=True)
df["duration"].fillna("Unkown_duration",inplace=True)

```

**Top 10 directors who have appeared in most movies or TV shows.**

```

import numpy as np
x=df.groupby("director")["title"].unique().sort_values(ascending=False)[:11].index
y=df.groupby("director")["title"].unique().sort_values(ascending=False)[:11].values

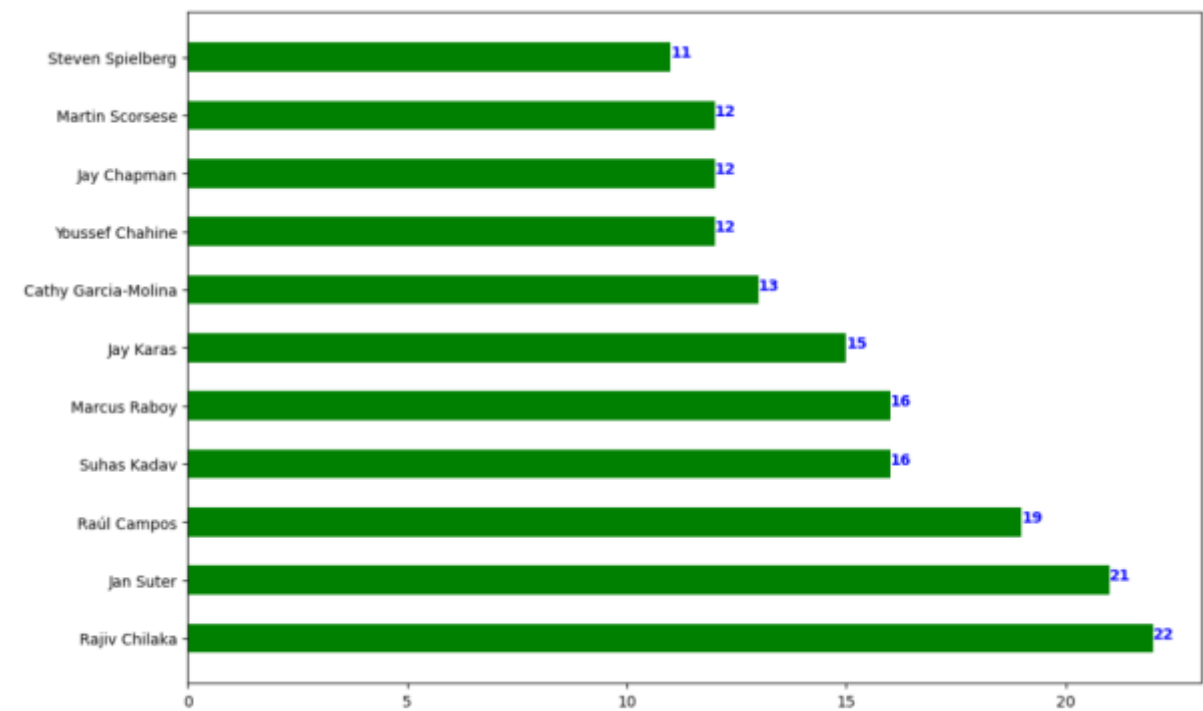
fig,ax=plt.subplots(figsize = (12,8))
width = 0.50

ax.barh(x,y,width,color="green")

for i,v in enumerate(y):
    ax.text(v,i,str(v),color='blue',fontweight = "bold")

plt.show()

```



#### Insights:

By observing the above data Rajiv Chilaka tops the list by directing 22 movies and followed by Jan Suter , Raul Campos ,Suhas Kadav Marcus Raboy and many more

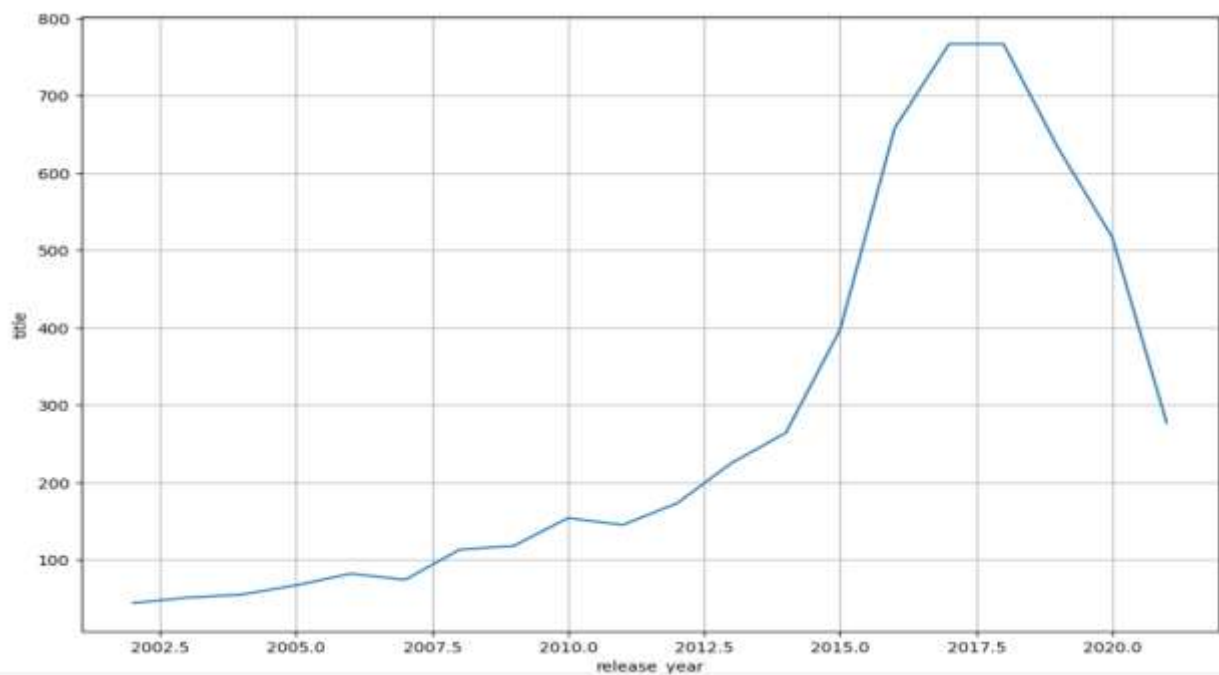
## How has the number of movies released per year changed over the last 20-30 years?

```
df_movies = df[df["type"] == "movies"]
df_sub_movies = df_movies.groupby("release_year")["title"].count().sort_values(ascending=False)[:20].reset_index()
df_sub_movies
```

	release_year	title
0	2018	767
1	2017	767
2	2016	658
3	2019	633
4	2020	517
5	2015	398

5	2015	398
6	2021	277
7	2014	264
8	2013	225
9	2012	173
10	2010	154
11	2011	145
12	2009	118
13	2008	113
14	2006	82
15	2007	74
16	2005	67
17	2004	55
18	2003	51
19	2002	44

```
plt.figure(figsize = (12,8))
sns.lineplot(x="release_year",y="title",data=df_sub_movies)
plt.grid()
plt.show()
```





**Insights:** By observing the above data we can conclude that in 2018 and 2017 years 767 movies were released in each year followed by 2016 and 2019 years with above 600 movies were released. By comparing with last 20 years there was a tremendous increase in movies releases, if we observe the line plot we can notice that in between 2015 to 2016 almost double number of movies were released in 2016 when comparing to 2015 there are so many reasons for this rise in movies. Because Netflix is an OTT platform which depends on internet data in between 2015 to 2016 internet boom happened in India Jio has been launched in 2016 with low internet data. This created a tremendous change in Digital services this is also one reasons. If internet data is cheaply available everyone can afford it also led to increase in memberships on Netflix platform many people started watching movies and TV-shows on Netflix with minimal internet charges. After 2018 is started decreasing if we observe in 2019 and 2020 there the graph is decreasing one of the reasons is due to covid-19 affect and increasing the internet data charges and membership prices.

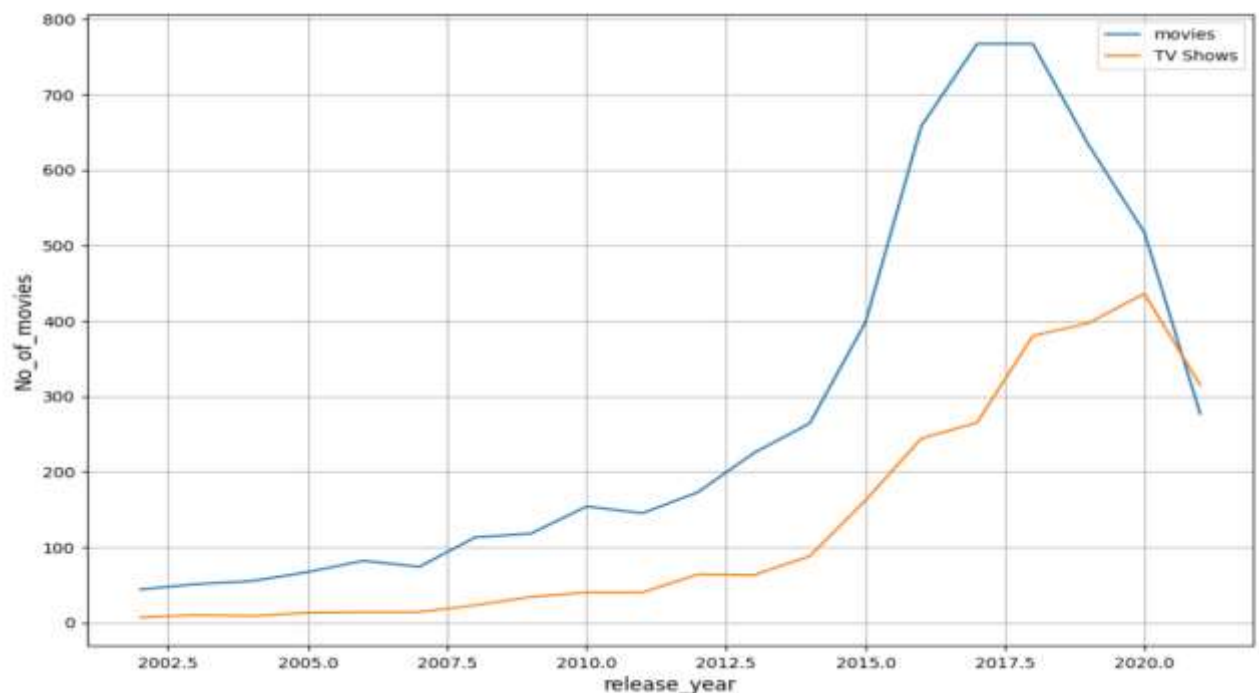
**Recommendations:** Lowering the membership prices can increase the customers. Netflix also focus on marketing which can create a big difference in sales on online platforms. It should also focus on investing on best IMBD rating movies and TV-shows instead of least ratings. Discounts can play a vital role in increasing memberships on particular year or monthly subscriptions providing discounts and yearly subscriptions help in increase in memberships.

## Comparison of tv shows vs. movies.

```
plt.figure(figsize = (12,8))

game1 = "movies"
sns.lineplot(x="release_year",y="title",data=df_sub_movies,label=game1)
plt.xlabel("release_year",fontsize = 12)
plt.ylabel("No_of_movies",fontsize = 12)

game2 = "TV Shows"
sns.lineplot(x="release_year",y="title",data=df_sub_shows,label = game2)
plt.legend(loc="upper right")
plt.grid()
plt.show()
```



**Insights:** By observing the line plot we can conclude that movies are dominating over TV-Shows for many years. There is a gradual increase in TV-shows but a sudden increase in movies between 2015 to 2016 because in 2011 to 2018 people used to follow old tradition like watching only movies and not aware of TV shows, there are very less TV shows were released comparing to Movies. One more reason is TV shows duration is long when comparing to movies so people preferred movies due to less duration. But in 2019 and 2020 there was rise in TV shows because due to covid effect and lockdown people got so much time to watch the TV shows this led to rise the demand for TV shows became high popular. But when people are turning out from movies to TV shows it led to decrease in movies after 2019 and 2020 both decreased.

**Recommendation:** Netflix need to focus on TV shows also which results in more profit due to its long duration of watching, we can observe that after 2019 and 2020 both decreased because people became busy after covid they are not spending too much time on the Netflix due to their other works but popularity for this online platforms is increasing so Netflix need to bring some new marketing strategies for attracting the audios need to implement the discounts policy on yearly subscriptions and need to encourage the Referrals with cash bonuses. Investing money on best IMDB rating movies and TV shows rather than normal movies this will give the good results for increasing the memberships for Netflix platform

## Top 10 Cast Names for different types of Movies

```
df.groupby("cast")["title"].nunique().sort_values(ascending=False)[:10]
```

cast	title
Anupam Kher	43
Shah Rukh Khan	35
Julie Tejawani	33
Naseeruddin Shah	32
Takahiro Sakurai	32
Rupa Bhimani	31
Akshay Kumar	30
Om Puri	30
Yuki Kaji	29
Amitabh Bachchan	28

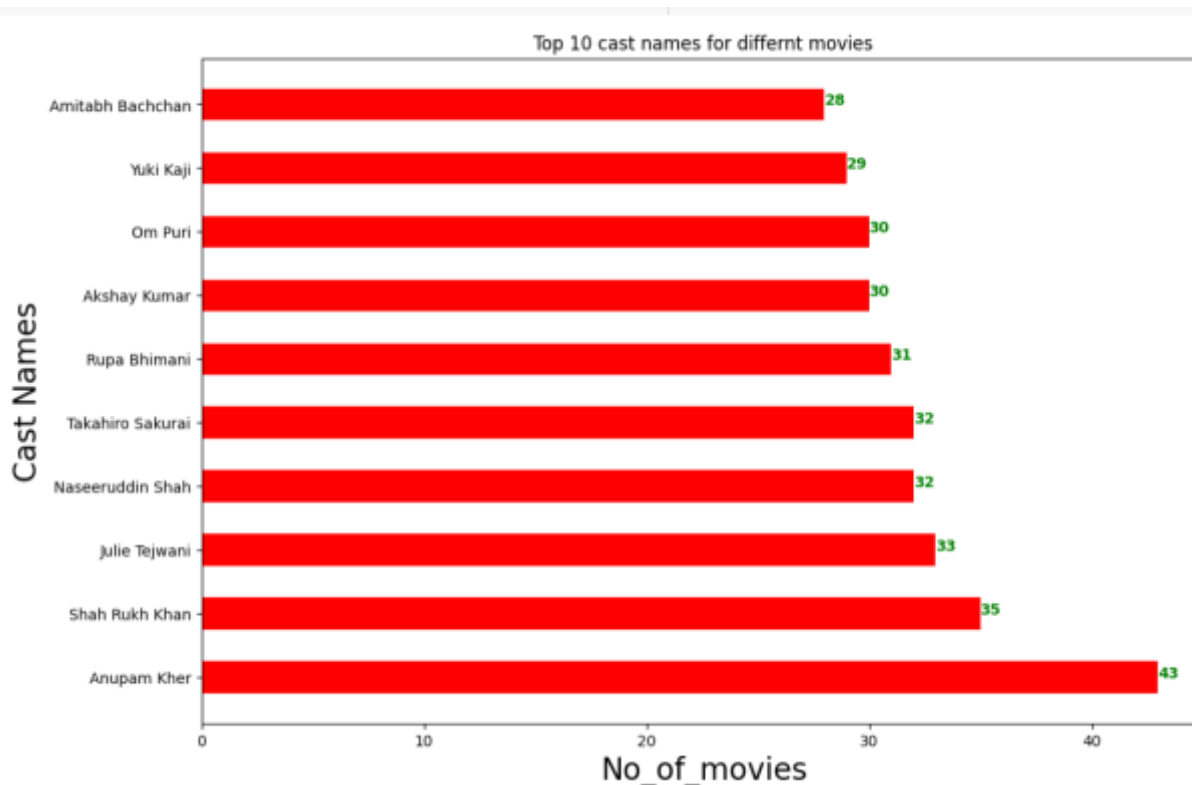
```
x = df.groupby("cast")["title"].nunique().sort_values(ascending=False)[:10].index
y = df.groupby("cast")["title"].nunique().sort_values(ascending=False)[:10].values

fig,ax=plt.subplots(figsize = (12,8))
width = 0.50

ax.barh(x,y,width,color="red")
plt.title("Top 10 cast names for differnt movies")
plt.ylabel("Cast Names",fontsize = (20))
plt.xlabel("No_of_movies",fontsize = (20))

for i,v in enumerate(y):
    ax.text(v,i,str(v),color='green',fontweight = "bold")

plt.show()
```



**Insights:** when we observe the Top 10 cast names for different movies Anupam kher holds the first position with 43 movies and second position owned by Shah Rukh khan with 35 movies followed by Julie Tejewani and Naseeruddin shah with 33 movies and finally Amitab Bachchan holds 10<sup>th</sup> position

**Recommendations:** Netflix need to focus on the famous cast movies which will result in attracting more number of customers to this platform. As we know that stardom plays a key role in watching a movie example it want to watch any movie first we will look the famous cast names in that particular movie many factors we will consider at first we will look whether our loved hero or heroine acted or casted in this movie or not. If we see Shah Rukh Khan is very famous actor if we take his movies failure percentage is very less when comparing to others. This strategy will help in increasing the Netflix profit

## Top 10 Cast Names for different types of TV Shows

```
df_sh1.groupby("cast")["title"].nunique().sort_values(ascending=False)[:10]
```

title	
cast	
Takahiro Sakurai	25
Yuki Kaji	19
Junichi Suwabe	17
Daisuke Ono	17
Al Kayano	17
Yuichi Nakamura	16
Yoshimasa Hosoya	15
Jun Fukuyama	15
David Attenborough	14
Kana Hanazawa	13

```

x = df_sh1.groupby("cast")["title"].nunique().sort_values(ascending=False)[:10].index
y = df_sh1.groupby("cast")["title"].nunique().sort_values(ascending=False)[:10].values

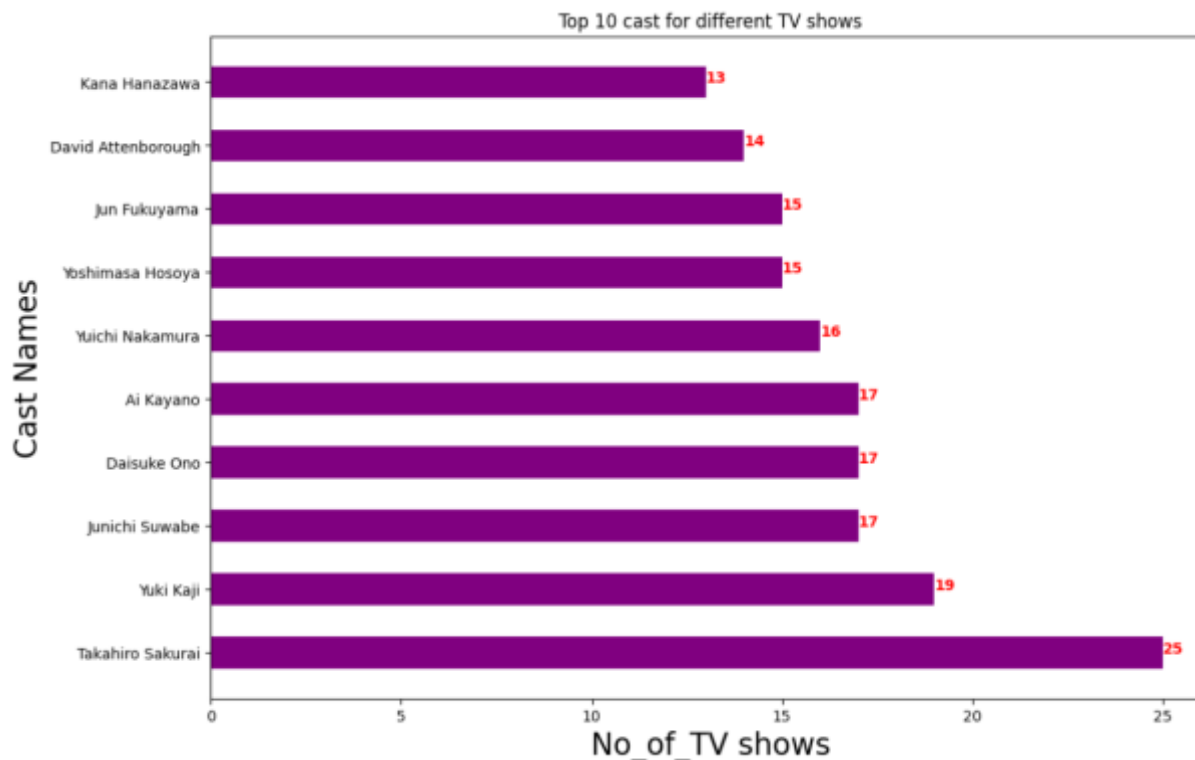
fig,ax=plt.subplots(figsize = (12,8))
width = 0.50

ax.barh(x,y,width,color="purple")
plt.title("Top 10 cast for different TV shows")
plt.ylabel("Cast Names",fontsize = (20))
plt.xlabel("No_of_TV shows",fontsize = (20))

for i,v in enumerate(y):
    ax.text(v,i,str(v),color='red',fontweight = "bold")

plt.show()

```



**Insights:** we move to the Tv shows Takahiro Sakurai holds the first position with 25 shows followed by yuki kaji with 19 shows and Junichi suwabe with 17 shows , Daisuke Ono with 17 shows Ai Kayano with 17 ,Yaichi Nakamura with 16, Yoshimasa with 15 shows many more

**Recommendations:** Netflix need to focus on the famous cast shows which will result in attracting more number of customers to this platform. As we know that stardom plays a key role in watching a movie example it want to watch any show first we will look the famous cast names in that particular movie many factors we will consider at first we will look whether our loved hero or heroine acted or casted in this movie or not. If we see Takahiro Sakurai is very famous actor if we take his shows failure percentage is very less when comparing to others. This strategy will help in increasing the Netflix profit

**Does Netflix has more focus on TV Shows than movies in recent years?**

## After 2015 Movies vs TV Shows

```
df['release_year'] = pd.to_numeric(df['release_year'])

recent_years = df[df['release_year'] >= 2015]

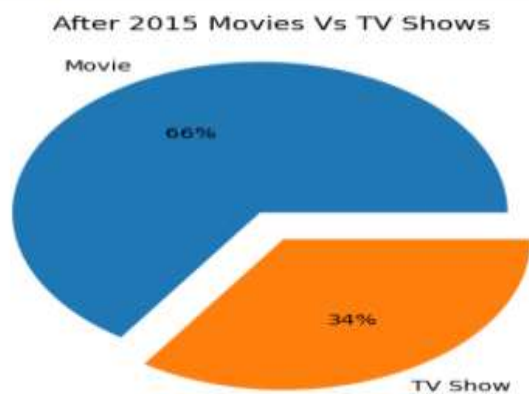
type_of_content = recent_years['type'].value_counts()

type_of_content
```

	count
type	
Movie	37575
TV Show	19778

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.pie(type_of_content,
        labels = type_of_content.index,
        autopct = "%.f%%",
        explode = [0.1,0.1]
    )
plt.title("After 2015 Movies Vs TV Shows")
plt.show()
```



## After 2018 movies vs TV Shows

```
df['release_year'] = pd.to_numeric(df['release_year'])

recent_years = df[df['release_year'] >= 2018]

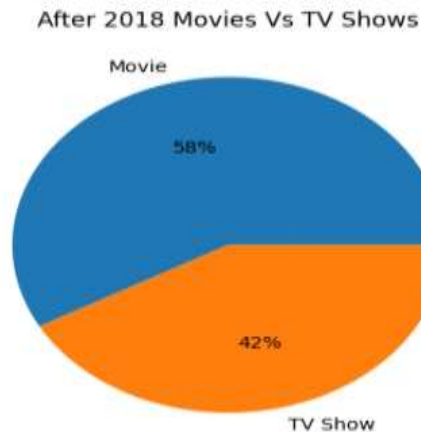
type_of_content = recent_years['type'].value_counts()

type_of_content
```

	count
type	
Movie	19605
TV Show	14156

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.pie(type_of_content,
        labels = type_of_content.index,
        autopct = "%.f%%"
    )
plt.title("After 2018 Movies Vs TV Shows")
plt.show()
```



**Insights:** By observing above data Netflix has more focus on TV Shows rather than movies. By taking the above pie charts into considerations, after 2015 piechart shows 66 % movies and 34% TV shows but when it comes to after 2018 years movies decreased to 58% and TV shows increased to 42% with above data. So it is clear that Netflix is more focusing on TV shows rather than movies.

**Recommendation:** Actually TV Shows take long duration when comparing to the movies ,some of the TV shows doesn't end in single season it will take 4 to 5 seasons to complete a one TV Show. One thing we need to notice here is even though TV shows duration is long now a days audios are prefer watching TV Shows because of they look interesting and crime thriller these attract more and more audios . If duration is long for particular TV Show in order to complete a that show audios will extend their memberships by paying extra money which makes profit business for Netflix. All seasons will not release at a single time this will built curiosity towards TV shows encouraging TV show will certainly benefits for the Netflix

## What type of content is available in different countries?

```
df_movie1 = df[df["type"]=="Movie"]
df_movie1 = df_movie1.groupby("country")["title"].nunique().sort_values(ascending=False)[:10]
df_movie1
```

country	title
United States	2751
India	962
United Kingdom	532
Canada	319
France	303
Germany	182
Spain	171
Japan	119
China	114
Mexico	111

```

x = df_movie1.index
y = df_movie1.values

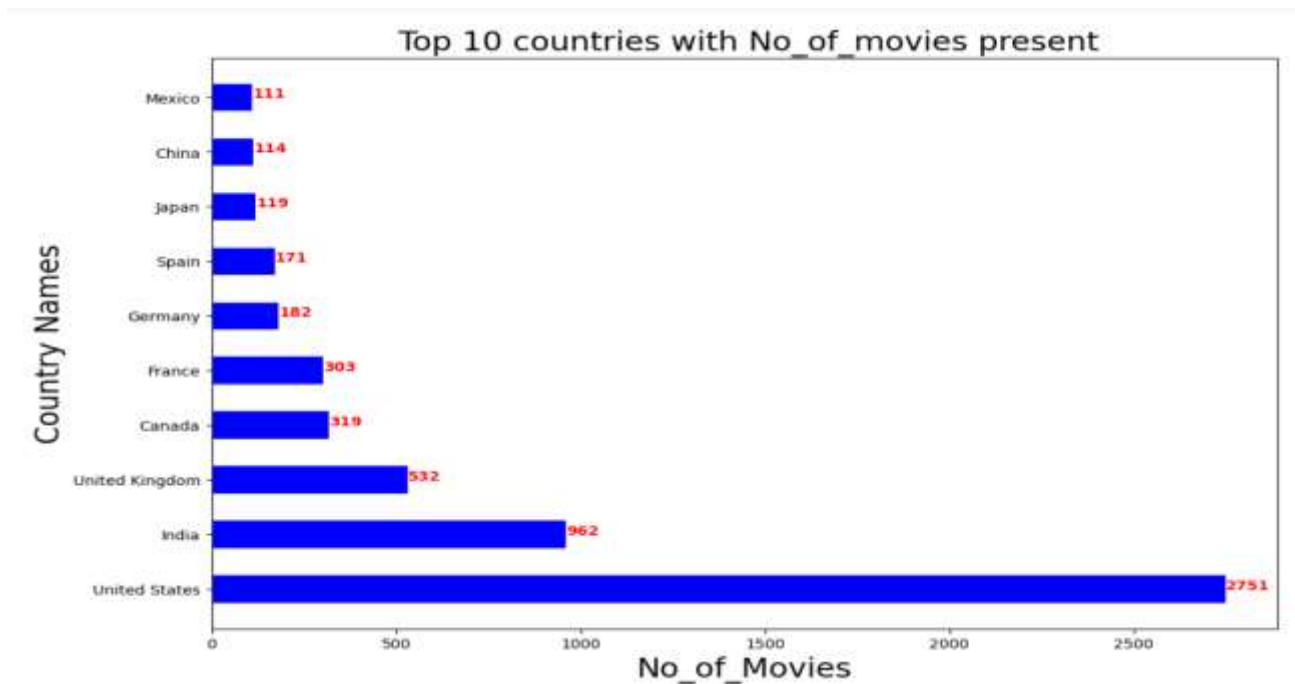
fig,ax=plt.subplots(figsize = (12,8))
width = 0.50

ax.barh(x,y,width,color="blue")
plt.title(" Top 10 countries with No_of_movies present",fontsize = (20))
plt.ylabel("Country Names",fontsize = (20))
plt.xlabel("No_of_Movies",fontsize = (20))

for i,v in enumerate(y):
    ax.text(v,i,str(v),color='red',fontweight = "bold")

plt.show()

```



**Insights:** By observing the above bar plot united states holds top position with 2751 movies followed by India with 962 movies the gap between first position and second is almost 3 times to second position. Third position holds by united kingdom with 532 movies ,Canada at fourth position with 319 movies and France at fifth position with 303 movies other countries are very less in number above hundred and below two hundred.

**Recommendations:** We can conclude that United country plays a vital role on Netflix platform because it was founded by American guys called Reed Hastings and Marc Randolph in Scotts Valley, California. Apart from United states Netflix is also growing in India in recent years. So it should focus more on India because due to its large population. Netflix can expect more and more audios in coming years from India but it should take care of some steps like maintaining minimal subscription price rates in India comparing to other countries. By understanding the nature of audios mindset Netflix should implement there plans in a proper way which can produce amazing results and good profits.

```
df_shows1 = df[df["type"]=="TV Show"]
df_shows1 = df_shows1.groupby("country")["title"].nunique().sort_values(ascending=False)[:10]
df_shows1
```

country	title
United States	938
United Kingdom	272
Japan	199
South Korea	170
Canada	126
France	90
India	84
Taiwan	70
Australia	66
Spain	61

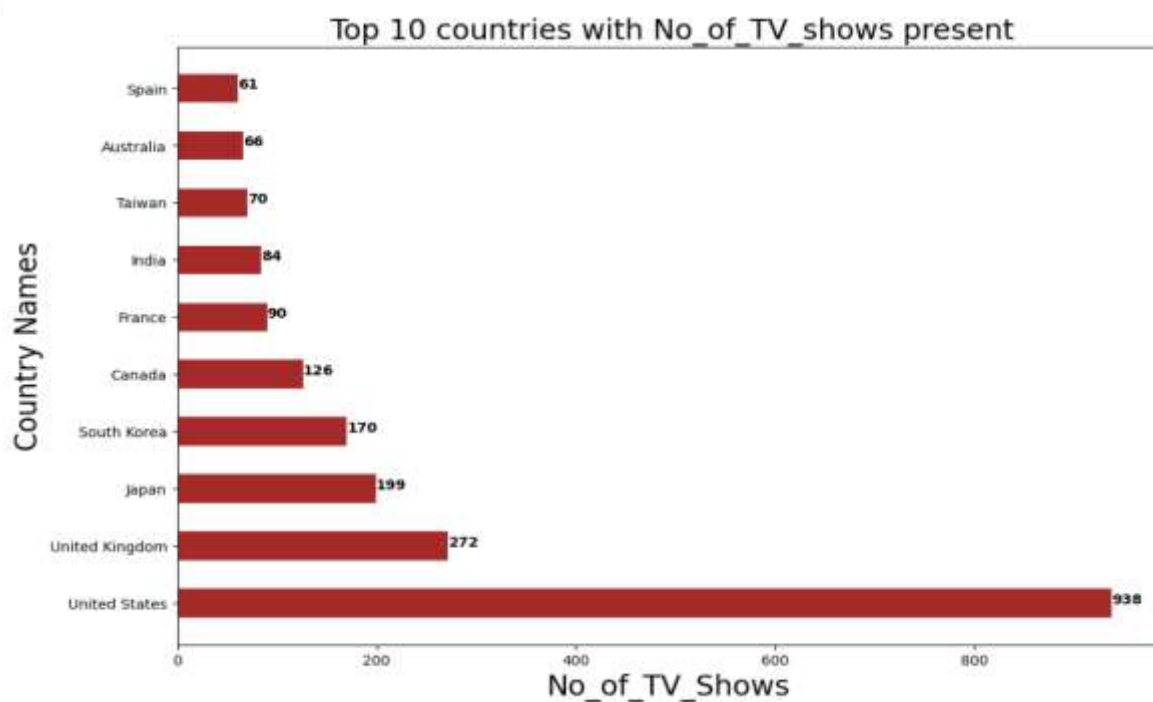
```
x = df_shows1.index
y = df_shows1.values

fig,ax=plt.subplots(figsize = (12,8))
width = 0.50

ax.barh(x,y,width,color="brown")
plt.title(" Top 10 countries with No_of_TV_shows present",fontsize = (20))
plt.ylabel("Country Names",fontsize = (20))
plt.xlabel("No_of_TV_Shows",fontsize = (20))

for i,v in enumerate(y):
    ax.text(v,i,str(v),color='black',fontweight = "bold")

plt.show()
```



**Insights:** By observing the above data united states holds the top position with 938 TV Shows followed by united



kingdom with 272 TV shows the gap between first position and second position is almost 4 times and Japan secured third position with 199 and south Korea with 170 and India is at seventh position other countries are less than 100 Tv shows .

**Recommendations:** We can conclude that United country plays a vital role on Netflix platform because it was founded by American guys called Reed Hastings and Marc Randolph in Scotts Valley, California. Apart from United states Netflix is also growing in India in recent years. So it should focus more on India because due to its large population. Netflix can expect more and more audios in coming years from India but it should take care of some steps like maintaining minimal subscription price rates in India comparing to other countries. By understanding the nature of audios mindset Netflix should implement there plans in a proper way which can produce amazing results and good profits.

## What is the best time to launch a TV show?

```
df['date_added'] = pd.to_datetime(df['date_added'])

df['Month'] = df['date_added'].dt.month
# Assuming 'tv_shows' is your DataFrame

tv_shows = df.query('type == "TV Show"')
# Assuming 'movies' is your DataFrame
movies = df.query('type == "Movie"')

tv_shows_monthly = tv_shows.groupby('Month')['show_id'].count()

movies_monthly = movies.groupby('Month')['show_id'].count()

best_tv_shows_month = tv_shows_monthly.idxmax()
best_movies_month = movies_monthly.idxmax()
```

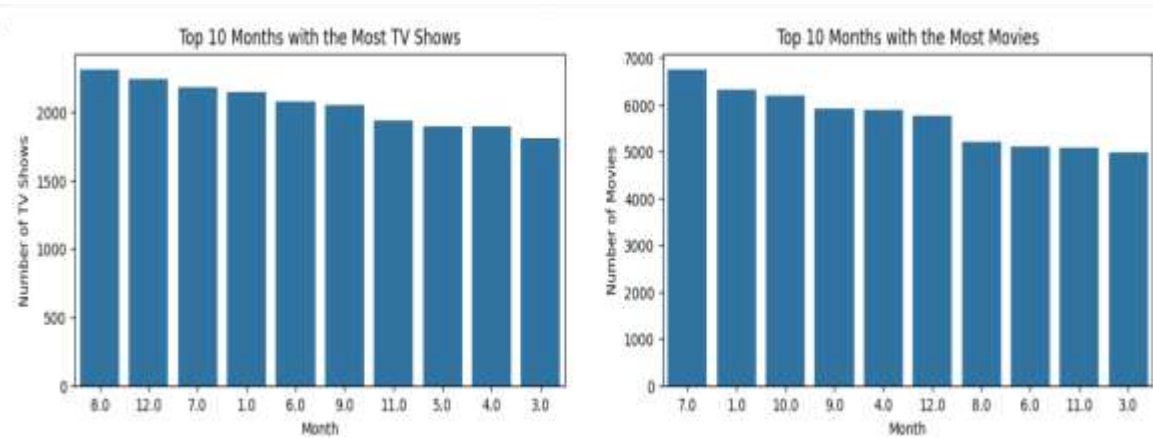
```
) import matplotlib.pyplot as plt
import seaborn as sns

tv_shows_monthly = tv_shows_monthly.sort_values(ascending=False).iloc[:10]
movies_monthly = movies_monthly.sort_values(ascending=False).iloc[:10]
plt.figure(figsize = (16,8))

plt.subplot(2,2,1)
sns.barplot(x=tv_shows_monthly.index,y=tv_shows_monthly.values,order=tv_shows_monthly.index)
plt.title('Top 10 Months with the Most TV Shows')
plt.xlabel('Month')
plt.ylabel('Number of TV Shows')

plt.subplot(2,2,2)
sns.barplot(x=movies_monthly.index,y=movies_monthly.values,order=movies_monthly.index)
plt.title('Top 10 Months with the Most Movies')

plt.xlabel('Month')
plt.ylabel('Number of Movies')
plt.show()
```



## Analysis on Weekly basis

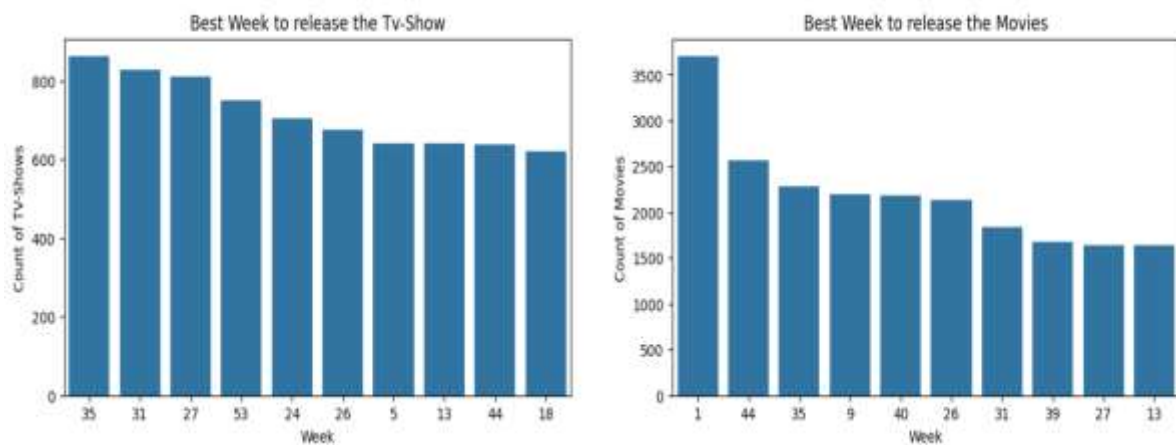
```
tv_shows_weekly = tv_shows_weekly.sort_values(ascending=False).iloc[:10]
movies_weekly = movies_weekly.sort_values(ascending=False).iloc[:10]

plt.figure(figsize = (16,9))
plt.subplot(2,2,1)
sns.barplot(x=tv_shows_weekly.index,y=tv_shows_weekly.values,order=tv_shows_weekly.index)

plt.title('Best Week to release the Tv-Show')
plt.xlabel('Week')
plt.ylabel('Count of TV-Shows')
plt.subplot(2,2,2)

sns.barplot(x=movies_weekly.index,y=movies_weekly.values,order=movies_weekly.index)

plt.title('Best Week to release the Movies')
plt.xlabel('Week')
plt.ylabel('Count of Movies')
plt.show()
```



## Insights

**Best time for TV Shows:** The analysis suggests that the best time to launch a TV show on Netflix is during the 27th week of the year. Additionally, the month of December stands out as a favourable period for TV show releases.

**Best time for Movies:** For movies, the best week to launch is the 1st week of the year, and the best month is July. These specific weeks and months are identified as peak times for movie releases. Movies are prominently released in weeks falling in July, early October, late February to early March, late June to early July, and late August to early September. This pattern suggests that movie production peaks around the beginning of summer, early fall, and late winter/early spring periods. Recommendations: Strategic Content Release: Plan content releases strategically based on insights about the best months for TV shows and movies. Aligning releases with peak months can maximize viewership and engagement. Promotions and Marketing: Implement marketing and promotional activities during the identified peak months to enhance visibility and attract a larger audience. Consider special campaigns or collaborations to boost content awareness

## Recommendations:

**Strategic Content Release:** Plan content releases strategically based on insights about the best months for TV shows and movies. Aligning releases with peak months can maximize viewership and engagement. Promotions and Marketing: Implement marketing and promotional activities during the identified peak months to enhance visibility and attract a larger audience. Consider special campaigns or collaborations to boost content awareness.

**Marketing Strategies:** Implement engagement strategies, such as interactive features, social media campaigns, or live events, during the identified best months and days. This can enhance the overall viewer experience. Continuous Monitoring: Regularly monitor viewership trends and update release strategies based on evolving audience preferences. Keep track of changing patterns to stay adaptable and responsive.

**Collaboration Opportunities:** Explore collaboration opportunities with influencers, other content creators, or events during the best months. Collaborative efforts can amplify the reach and impact of content releases. By incorporating these recommendations, Netflix can optimize its content release strategy, improve audience engagement, and maintain a dynamic and successful platform throughout the year.

## Different types of content present on Netflix ( analysis based on rating)

```
def age_group(x):
    if x in ['TV-Y', 'TV-G', 'G']:
        return 'Little Kids'
    elif x in ['TV-Y7', 'TV-Y7-FV', 'TV-PG', 'PG']:
        return 'Older Kids'
    elif x in ['PG-13', 'TV-14']:
        return 'Teens'
    else:
        return 'Adults'

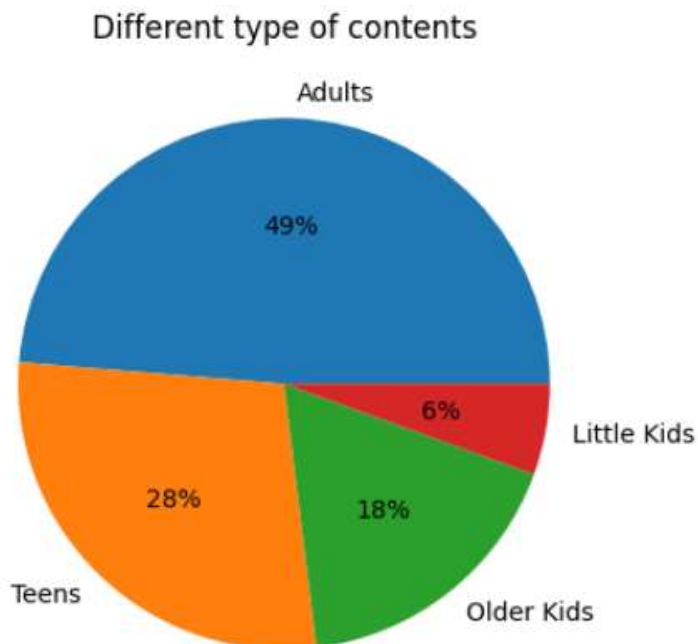
df_sub1 = df["content_type"].value_counts()
df_sub1
```

	count
content_type	
Adults	43529
Teens	25251
Older Kids	15650
Little Kids	4952

dtype: int64

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.pie(df_sub1,
        labels = df_sub1.index,
        autopct = "%.f%%"
        )
plt.title("Different type of contents")
plt.show()
```



### Insights:

By observing the above data adult type of content is dominating on Netflix when comparing to the other types. Almost half type of content is related to adults and followed by teens content with 28 percentage, older kids with 18 percentage least is 6 percentage content is related to Little kids

### Recommendations:

Netflix need to focus on Adult type of content where most of audios or viewers are belong to adults category. According to some analysis they prefer adults type of content and Horror movies this belong to adult category. When it comes to Teens category they prefer love and romantic type of movies and shows they may also like horrors movies and shows. Few older watch normal type of content like action thriller movies and shows little kids category prefer anime movies and kids type of animation movies and TV shows . Netflix need to their marketing by taking the above points into considerations which helps in better results.