Importing Required Liabraries

In [93]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Reading The Data_set

In [94]:
```
!gdown "https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/000/940/origi
```

```
Downloading...
From: https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/000/940/origin
al/netflix.csv (https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/000/
940/original/netflix.csv)
To: C:\Users\Satyam\netflix.csv

  0%|          | 0.00/3.40M [00:00<?, ?B/s]
 31%|###       | 1.05M/3.40M [00:00<00:00, 9.38MB/s]
 93%|#########2| 3.15M/3.40M [00:00<00:00, 15.1MB/s]
100%|##########| 3.40M/3.40M [00:00<00:00, 14.2MB/s]
```

In [95]:
```
df =pd.read_csv("netflix.csv")
```

Initial few basic Steps

In [96]:
```
df.head(10)
```

Out[96]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	durat
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 r
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	Seas
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	Seas
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	Seas
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	Seas
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	Seas
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	NaN	September 24, 2021	2021	PG	91 r
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 r
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...	United Kingdom	September 24, 2021	2021	TV-14	Seas
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG-13	104 r

In [97]:
```
df.shape
```

Out[97]:
```
(8807, 12)
```

In [98]:
```
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 [99]:
```
df.describe()
```

Out[99]:

	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

In [100]:
```
df.describe(include="object")
```

Out[100]:

	show_id	type	title	director	cast	country	date_added	rating	duration	listed
count	8807	8807	8807	6173	7982	7976	8797	8803	8804	8
unique	8807	2	8807	4528	7692	748	1767	17	220	5
top	s1	Movie	Dick Johnson Is Dead	Rajiv Chilaka	David Attenborough	United States	January 1, 2020	TV-MA	1 Season	Dram Internatic Mov
freq	1	6131	1	19	19	2818	109	3207	1793	3

After Seeing the data set i observed that the some of column have nested value so first we need to unnest the value of that column

Unnesting the columns value

In [101]:
```
df["cast"]=df['cast'].apply(lambda x: str(x).split(', ')).tolist()
df["director"] = df['director'].apply(lambda x: str(x).split(', ')).tolist()
df["country"] = df['country'].apply(lambda x: str(x).split(', ')).tolist()
df["listed_in"] = df['listed_in'].apply(lambda x: str(x).split(', ')).tolist()
```

In [102]:
```
df=df.explode("cast")
df=df.explode("director")
df=df.explode("country")
df=df.explode("listed_in")
```

In [103]:
```
df.head(10)
```

Out[103]:

	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	Do
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	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	
1	s2	TV Show	Blood & Water	nan	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	T
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	Khosi Ngema	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	T
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	Gail Mabalane	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	T

In [104]:
```
df.shape
```

Out[104]:
```
(201991, 12)
```

In [105]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 201991 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count   Dtype 
---  ------        --------------   ----- 
 0   show_id       201991 non-null  object
 1   type          201991 non-null  object
 2   title         201991 non-null  object
 3   director      201991 non-null  object
 4   cast          201991 non-null  object
 5   country       201991 non-null  object
 6   date_added    201833 non-null  object
 7   release_year  201991 non-null  int64 
 8   rating        201924 non-null  object
 9   duration      201988 non-null  object
 10  listed_in     201991 non-null  object
 11  description   201991 non-null  object
dtypes: int64(1), object(11)
memory usage: 20.0+ MB
```

In [106]:
```
df.replace("nan",np.nan,inplace= True)
```

In [107]:
```
df
```

Out[107]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duratio
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 mi
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	Season
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	Season
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	Season
1	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	Season
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 mi
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	111 mi
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 mi
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 mi
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	111 mi

201991 rows × 12 columns

In [108]:
```
#  -- description column is not useful so we need to drop this column
df.drop(columns=["description"],inplace=True)
```

In [109]:
```
# duration column has two types of values one is for tv show which is in season
# and another is for movies which is in min so wee need to split them and take numer
df["duration"]=df["duration"].apply(lambda x:str(x).split()[0])
```

In [110]:
```
df
```

Out[110]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duratio
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	9
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	September 24, 2021	2021	TV-MA	
1	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	September 24, 2021	2021	TV-MA	
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	March 2, 2019	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	March 2, 2019	2015	TV-14	11

201991 rows × 11 columns

In [111]:
```
# date_added column data type is object we have to change the data type of column is
df["date_added"]=pd.to_datetime(df["date_added"])
```

In [112]:
```
df
```

Out[112]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duratio
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	9
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	
1	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2019-03-02	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2019-03-02	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	11

201991 rows × 11 columns

In [113]:
```
import datetime as dt
```

In [114]:
```
df["month"]= df["date_added"].dt.month
```

In [115]:
```
df["year"]= df["date_added"].dt.year
```

In [116]:
```
df["week_day_name"] =df["date_added"].dt.day_name()
```

In [117]:
```
df
```

Out[117]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duratio
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13	9
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata	South Africa	2021-09-24	2021	TV-MA	
1	s2	TV Show	Blood & Water	NaN	Khosi Ngema	South Africa	2021-09-24	2021	TV-MA	
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2019-03-02	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2019-03-02	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	11
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2019-03-02	2015	TV-14	11

201991 rows × 14 columns

In [118]:
```
# After extracting the month,date, year and week_day_name from this column
# date_added columnn is no longer is usefull so we have to drop this column
df.drop(columns=["date_added"],inplace=True)
```

In [119]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 201991 entries, 0 to 8806
Data columns (total 13 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   show_id        201991 non-null  object 
 1   type           201991 non-null  object 
 2   title          201991 non-null  object 
 3   director       151348 non-null  object 
 4   cast           199845 non-null  object 
 5   country        190094 non-null  object 
 6   release_year   201991 non-null  int64  
 7   rating         201924 non-null  object 
 8   duration       201991 non-null  object 
 9   listed_in      201991 non-null  object 
 10  month          201833 non-null  float64
 11  year           201833 non-null  float64
 12  week_day_name  201833 non-null  object 
dtypes: float64(2), int64(1), object(10)
memory usage: 21.6+ MB
```

After unnesting the required column i observed that some of column has missing value so we have to fill those missing value

Handling of missing values

In [120]:
```
df["cast"].mode()[0]
```

Out[120]: 'Liam Neeson'

In [121]:
```
df["director"].mode()[0]
```

Out[121]: 'Martin Scorsese'

In [122]:
```
df["listed_in"].mode()[0]
```

Out[122]: 'Dramas'

In [123]:
```
df["country"].mode()[0]
```

Out[123]: 'United States'

In [124]:
```
df["rating"].mode()[0]
```

Out[124]: 'TV-MA'

In [125]:
```
df["month"].mode()[0]
```

Out[125]: 7.0

In [126]:
```
df["week_day_name"].mode()[0]
```

Out[126]:
```
'Friday'
```

In [127]:
```
df["year"].mode()[0]
```

Out[127]:
```
2019.0
```

In [128]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 201991 entries, 0 to 8806
Data columns (total 13 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   show_id        201991 non-null  object 
 1   type           201991 non-null  object 
 2   title          201991 non-null  object 
 3   director       151348 non-null  object 
 4   cast           199845 non-null  object 
 5   country        190094 non-null  object 
 6   release_year   201991 non-null  int64  
 7   rating         201924 non-null  object 
 8   duration       201991 non-null  object 
 9   listed_in      201991 non-null  object 
 10  month          201833 non-null  float64
 11  year           201833 non-null  float64
 12  week_day_name  201833 non-null  object 
dtypes: float64(2), int64(1), object(10)
memory usage: 21.6+ MB
```

In [129]:
```
df["cast"].fillna(df["cast"].mode()[0],inplace=True)
```

In [130]:
```
df["director"].fillna(df["director"].mode()[0],inplace=True)
```

In [131]:
```
df["listed_in"].fillna(df["listed_in"].mode()[0],inplace=True)
df["country"].fillna(df["country"].mode()[0],inplace=True)
df["rating"].fillna(df["rating"].mode()[0],inplace=True)
df["month"].fillna(df["month"].mode()[0],inplace=True)
df["year"].fillna(df["year"].mode()[0],inplace=True)
df["week_day_name"].fillna(df["week_day_name"].mode()[0],inplace=True)
```

In [132]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 201991 entries, 0 to 8806
Data columns (total 13 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   show_id        201991 non-null  object 
 1   type           201991 non-null  object 
 2   title          201991 non-null  object 
 3   director       201991 non-null  object 
 4   cast           201991 non-null  object 
 5   country        201991 non-null  object 
 6   release_year   201991 non-null  int64  
 7   rating         201991 non-null  object 
 8   duration       201991 non-null  object 
 9   listed_in      201991 non-null  object 
 10  month          201991 non-null  float64
 11  year           201991 non-null  float64
 12  week_day_name  201991 non-null  object 
dtypes: float64(2), int64(1), object(10)
memory usage: 21.6+ MB
```

In [133]:
```
df.head()
```

Out[133]:

	show_id	type	title	director	cast	country	release_year	rating	duration	listed_in	n
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Liam Neeson	United States	2020	PG-13	90	Documentaries	
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	International TV Shows	
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Dramas	
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Mysteries	
1	s2	TV Show	Blood & Water	Martin Scorsese	Khosi Ngema	South Africa	2021	TV-MA	2	International TV Shows	

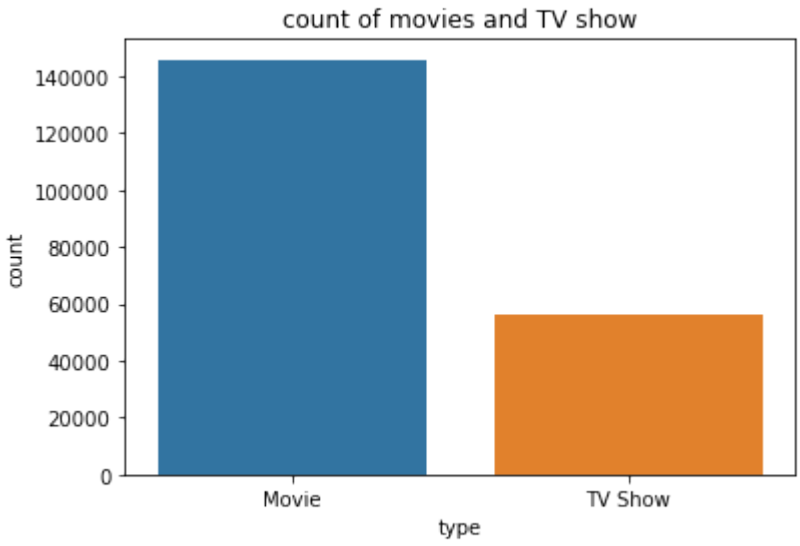
Analysis 1- How much movies and TV shows listed on Netflix

In [134]:
```
df.groupby(["type"])["title"].count()
```

Out[134]:
```
type
Movie      145843
TV Show     56148
Name: title, dtype: int64
```

In [135]:
```
sns.countplot(data=df,x="type")
plt.title("count of movies and TV show")
```

Out[135]:
```
Text(0.5, 1.0, 'count of movies and TV show')
```



Insight- After analysing the data i found that more number of movies listed on Netflix as compared to TV Show

Analysis-2 Number of movies and tv shows added on netflix in recent years

In [136]:
```
df["year"].max()
```

Out[136]:
```
2021.0
```

In [137]:
```
df["month"].min()
```

Out[137]:
```
1.0
```

In [138]:
```
df["year"].min()
```

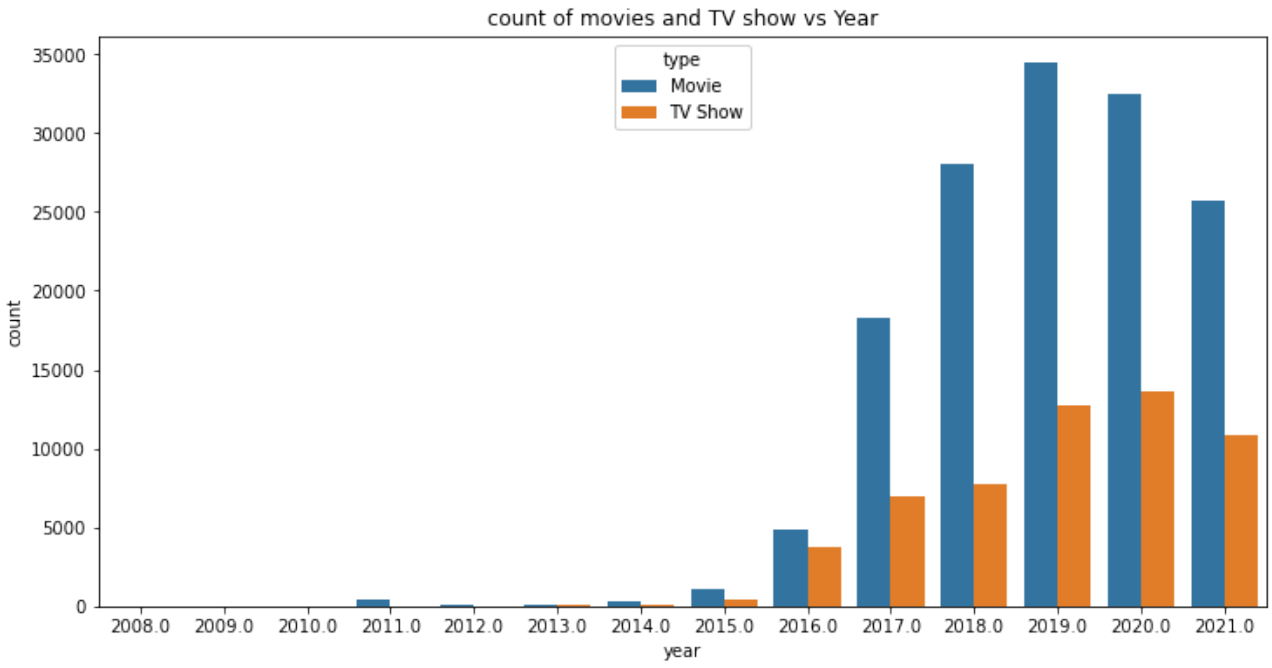
Out[138]:
```
2008.0
```

In [139]:
```
df.groupby(["year","type"])["type"].value_counts()
```

Out[139]:
```
year    type     type
2008.0  Movie    Movie         18
        TV Show  TV Show        1
2009.0  Movie    Movie         30
2010.0  Movie    Movie         20
2011.0  Movie    Movie        438
2012.0  Movie    Movie         36
2013.0  Movie    Movie         75
        TV Show  TV Show      132
2014.0  Movie    Movie        343
        TV Show  TV Show      107
2015.0  Movie    Movie       1125
        TV Show  TV Show      435
2016.0  Movie    Movie       4858
        TV Show  TV Show      3711
2017.0  Movie    Movie      18252
        TV Show  TV Show      6957
2018.0  Movie    Movie      28049
        TV Show  TV Show      7735
2019.0  Movie    Movie      34446
        TV Show  TV Show     12682
2020.0  Movie    Movie      32462
        TV Show  TV Show     13563
2021.0  Movie    Movie      25691
        TV Show  TV Show     10825
Name: type, dtype: int64
```

In [140]:
```
plt.figure(figsize=(12,6))
sns.countplot(data=df,x="year",hue="type")
plt.title("count of movies and TV show vs Year")
```

Out[140]:
```
Text(0.5, 1.0, 'count of movies and TV show vs Year')
```



Insight- 1-After analysing the data i found that First content had added on Netflix in Jan 2018.

2-In last 5 to 6 years Netflix had more focus on movies.

Analysis-3 Best time to launch a TV show Movies on Netflix

In [ ]:

In [141]:
```
df_Tv_show= df.loc[df["type"]=="TV Show"]
```

In [142]:
```
df_Tv_show
```

Out[142]:

	show_id	type	title	director	cast	country	release_year	rating	duration	listed_in
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	International TV Shows
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Dramas
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Mysteries
1	s2	TV Show	Blood & Water	Martin Scorsese	Khosi Ngema	South Africa	2021	TV-MA	2	International TV Shows
1	s2	TV Show	Blood & Water	Martin Scorsese	Khosi Ngema	South Africa	2021	TV-MA	2	TV Dramas
...	...	...	...	...	...	...	...	...	...	...
8800	s8801	TV Show	Zindagi Gulzar Hai	Martin Scorsese	Hina Khawaja Bayat	Pakistan	2012	TV-PG	1	Romantic TV Shows
8800	s8801	TV Show	Zindagi Gulzar Hai	Martin Scorsese	Hina Khawaja Bayat	Pakistan	2012	TV-PG	1	TV Dramas
8803	s8804	TV Show	Zombie Dumb	Martin Scorsese	Liam Neeson	United States	2018	TV-Y7	2	Kids' TV
8803	s8804	TV Show	Zombie Dumb	Martin Scorsese	Liam Neeson	United States	2018	TV-Y7	2	Korean TV Shows
8803	s8804	TV Show	Zombie Dumb	Martin Scorsese	Liam Neeson	United States	2018	TV-Y7	2	TV Comedies

56148 rows × 13 columns

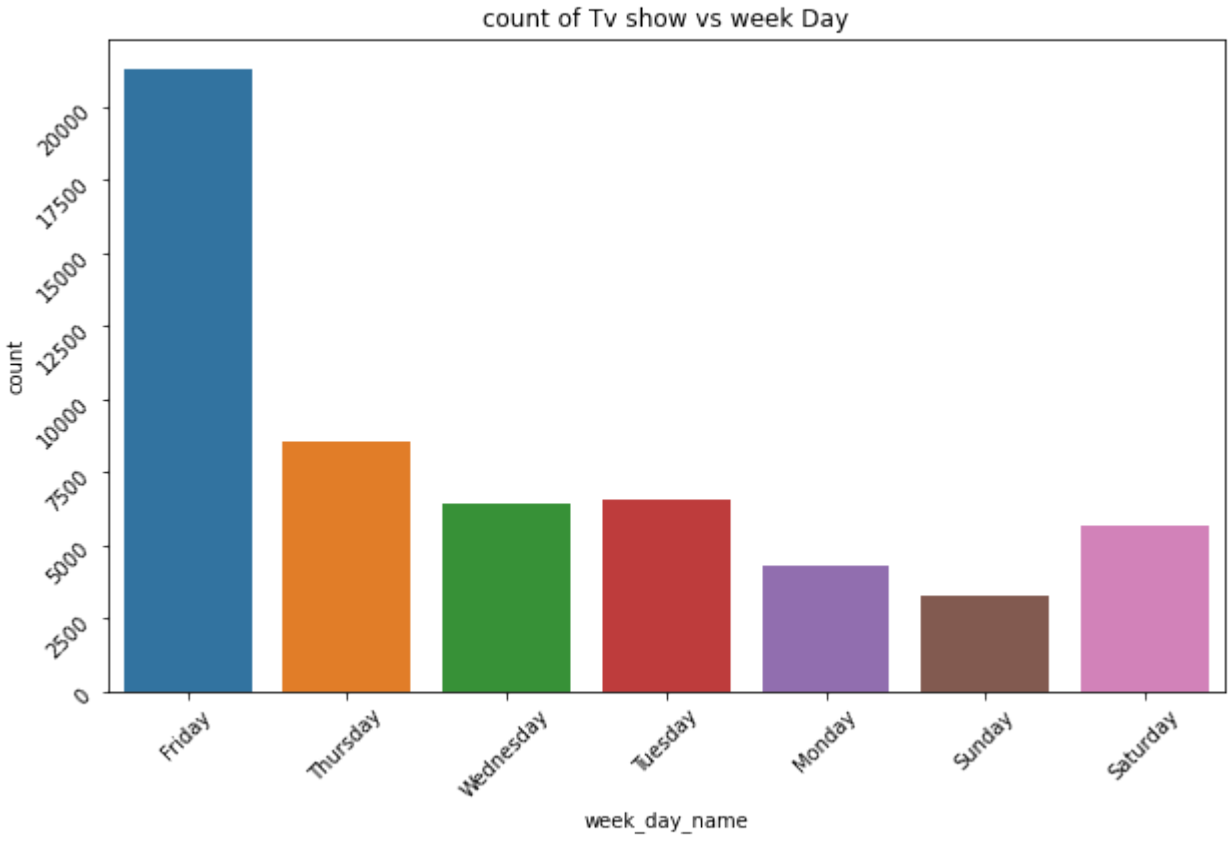
In [143]:
```
df_Tv_show.groupby("week_day_name")["title"].count()
```

Out[143]:
```
week_day_name
Friday       21284
Monday        4306
Saturday      5664
Sunday        3302
Thursday      8554
Tuesday       6604
Wednesday     6434
Name: title, dtype: int64
```

In [144]:
```
plt.figure(figsize=(10,6))
sns.countplot(data=df_Tv_show,x="week_day_name")
plt.xticks(rotation=45)

plt.yticks(rotation=45)
plt.title("count of Tv show vs week Day")

plt.show()
```



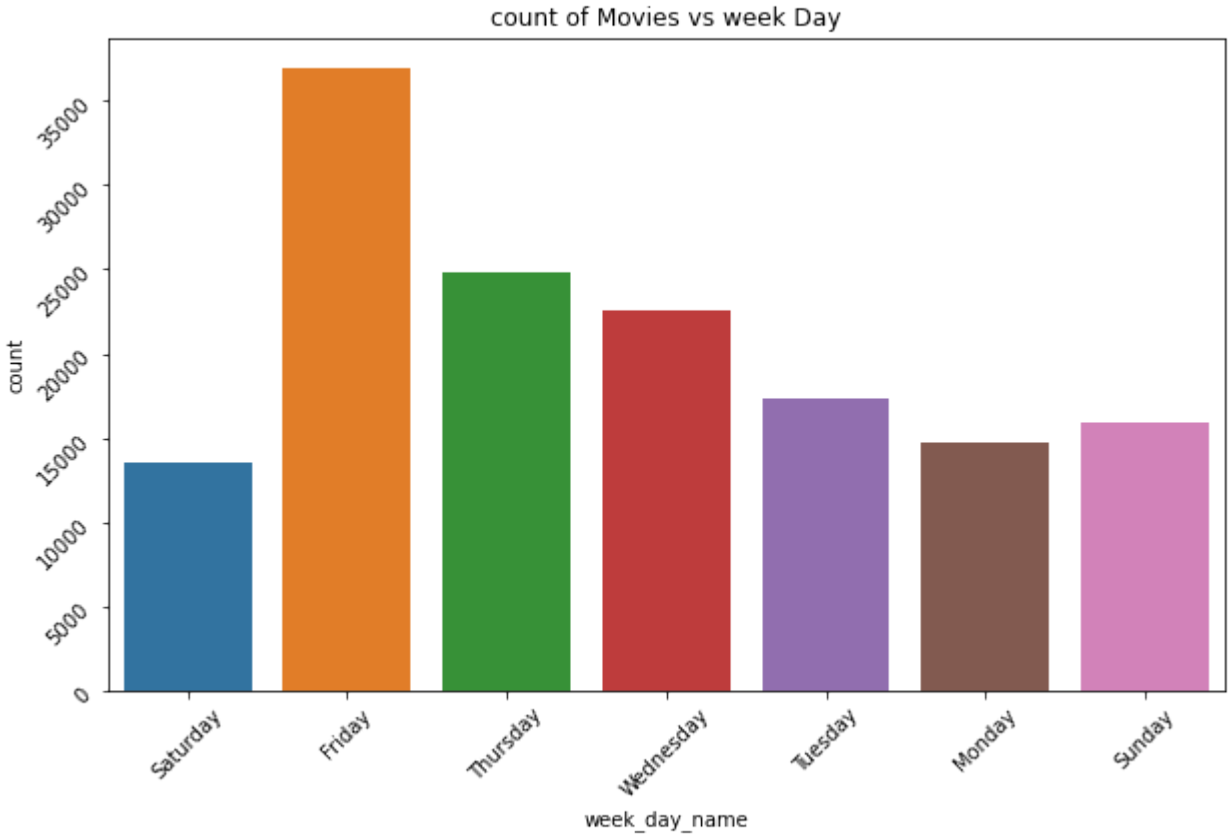
In [145]:
```
df_Movies= df.loc[df["type"]=="Movie"]
```

In [146]:
```
df_Movies.groupby(["week_day_name"])["title"].count()
```

Out[146]:
```
week_day_name
Friday       36899
Monday       14681
Saturday     13588
Sunday       15944
Thursday     24860
Tuesday      17371
Wednesday    22500
Name: title, dtype: int64
```

In [147]:
```
plt.figure(figsize=(10,6))
sns.countplot(data=df_Movies,x="week_day_name")
plt.xticks(rotation=45)

plt.yticks(rotation=45)
plt.title("count of Movies vs week Day")
plt.show()
```



Insight - After analysing the data i found that the More content added on Netflix on Friday as compared to other days

Analysis-4 Number of Movies released per year changed over the last 20-30 years

In [148]:
```
recent_year_data= df.loc[(df["release_year"]>1992) & (df["type"]=="Movie")]
recent_year_data
```

Out[148]:

	show_id	type	title	director	cast	country	release_year	rating	duration	list
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Liam Neeson	United States	2020	PG-13	90	Documer
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	United States	2021	PG	91	Chil Family M
6	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Vanessa Hudgens	United States	2021	PG	91	Chil Family M
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Kimiko Glenn	United States	2021	PG	91	Chil Family M
6	s7	Movie	My Little Pony: A New Generation	José Luis Ucha	Kimiko Glenn	United States	2021	PG	91	Chil Family M
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2015	TV-14	111	Interna M
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2015	TV-14	111	M Mu
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	D
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Interna M
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	M Mu

138272 rows × 13 columns

In [149]:
```
df5= recent_year_data.groupby(["release_year"])["title"].count().reset_index()
```

In [150]:
```
df5.rename(columns={"title":"count"},inplace= True)
```

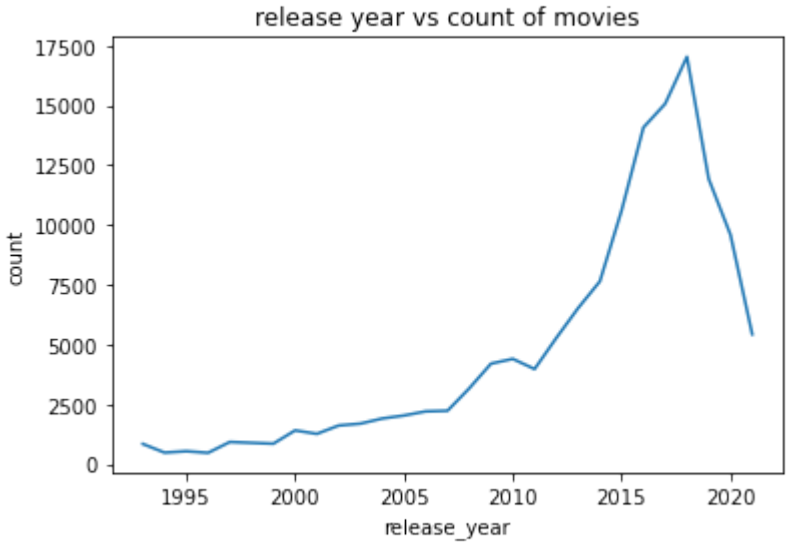
In [151]:
```
df5
```

Out[151]:

	release_year	count
0	1993	844
1	1994	480
2	1995	539
3	1996	471
4	1997	925
5	1998	889
6	1999	854
7	2000	1409
8	2001	1265
9	2002	1613
10	2003	1692
11	2004	1906
12	2005	2032
13	2006	2205
14	2007	2231
15	2008	3169
16	2009	4201
17	2010	4400
18	2011	3974
19	2012	5277
20	2013	6522
21	2014	7642
22	2015	10612
23	2016	14075
24	2017	15069
25	2018	17033
26	2019	11926
27	2020	9590
28	2021	5427

In [152]:
```
sns.lineplot(x= df5["release_year"],y = df5["count"])
plt.title("release year vs count of movies")
```

Out[152]:
```
Text(0.5, 1.0, 'release year vs count of movies')
```



Insight - After analysing the data i found that from 1993 till 2020 count of movie release was increasing but in last 2 years it starts decreasing

Analysis-5 Who is most famous director on Netflix

In [153]:
```
df.groupby("director")["title"].count().head(10)
```

Out[153]:
```
director
A. L. Vijay          42
A. Raajdheep         10
A. Salaam            30
A.R. Murugadoss      30
Aadish Keluskar      12
Aamir Bashir         12
Aamir Khan           14
Aanand Rai           30
Aaron Burns          12
Aaron Hancox          2
Name: title, dtype: int64
```

Insight- A.L. Vijay is most famous director on Netflix

Analysis-6 Who is most famous Actor on Netflix

In [154]:
```
df.groupby(["cast"])["title"].count().sort_values(ascending=False).head(10)
```

Out[154]:
```
cast
Liam Neeson          2307
Alfred Molina         160
John Krasinski        139
Salma Hayek           130
Frank Langella        128
Anupam Kher           127
John Rhys-Davies      125
Shah Rukh Khan        108
Naseeruddin Shah      106
Radhika Apte          104
Name: title, dtype: int64
```

Insight- Liam Neeson is most Famous Actor on Netflix.

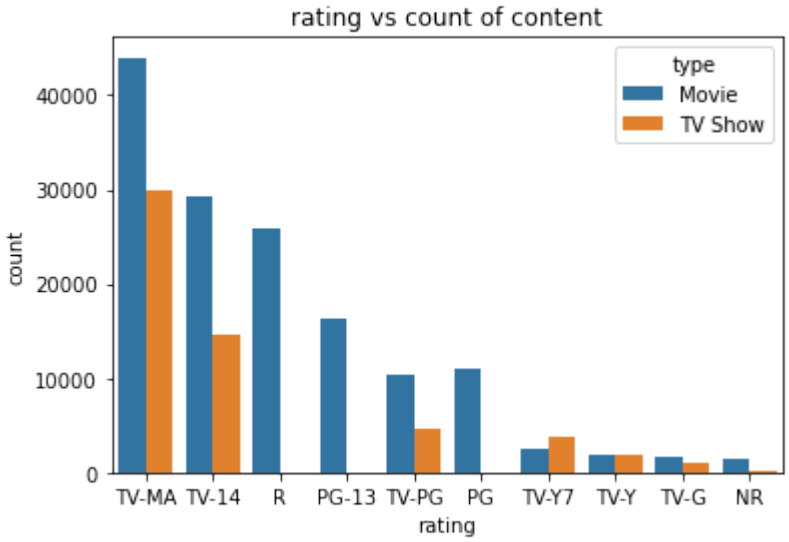
Analysis-7 Rating wise content on Netflix

In [155]:
```
df.groupby("rating")["title"].count().sort_values(ascending=False).head(10)
```

Out[155]:
```
rating
TV-MA    73934
TV-14    43931
R        25860
PG-13    16246
TV-PG    14926
PG       10919
TV-Y7     6304
TV-Y      3665
TV-G      2779
NR        1573
Name: title, dtype: int64
```

In [156]:
```
sns.countplot(data=df,x="rating",order=df["rating"].value_counts().index[0:10],hue="
plt.title("rating vs count of content")
```

Out[156]:
```
Text(0.5, 1.0, 'rating vs count of content')
```



Insight- After analysing the data on Netflix top 3 content rating is TV-MA, TV-14,R.

Analysis-8 Popular Actor and Director Combo on Netflix

In [157]:
```
df
```

Out[157]:

	show_id	type	title	director	cast	country	release_year	rating	duration	liste
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Liam Neeson	United States	2020	PG-13	90	Documenta
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	Internati TV Sh
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Dra
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Myste
1	s2	TV Show	Blood & Water	Martin Scorsese	Khosi Ngema	South Africa	2021	TV-MA	2	Internati TV Sh
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2015	TV-14	111	Internati Mc
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2015	TV-14	111	Mus Mus
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Dra
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Internati Mc
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Mus Mus

201991 rows × 13 columns

In [158]:
```
df2 = df.groupby(["cast","director"])["title"].nunique().sort_values(ascending=False
```

In [159]:
```
df2.rename(columns={"title":"count_title"},inplace=True)
```

In [160]:
```
df2
```

Out[160]:

	cast	director	count_title
0	Liam Neeson	Martin Scorsese	353
1	Takahiro Sakurai	Martin Scorsese	24
2	Julie Tejwani	Rajiv Chilaka	19
3	Rajesh Kava	Rajiv Chilaka	19
4	Jigna Bhardwaj	Rajiv Chilaka	18
...	...	...	...
62705	Hilliary Begley	Anne Fletcher	1
62706	Him Law	Teddy Chan	1
62707	Hima Singh	Jatla Siddartha	1
62708	Himani Shivpuri	K.C. Bokadia	1
62709	Ṣọpẹ́ Dìrísù	Remi Weekes	1

62710 rows × 3 columns

Insight- The most Famous actor director combos on Netflix is Liam Neeson and Martin Scorsese

Analysis-9 Distribution of Movies and TV show Across Countries on Netflix

In [161]:
```
df
```

Out[161]:

	show_id	type	title	director	cast	country	release_year	rating	duration	liste
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Liam Neeson	United States	2020	PG-13	90	Documenta
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	Internati TV Sh
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Dra
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Myste
1	s2	TV Show	Blood & Water	Martin Scorsese	Khosi Ngema	South Africa	2021	TV-MA	2	Internati TV Sh
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2015	TV-14	111	Internati Mc
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2015	TV-14	111	Mus Mus
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Dra
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Internati Mc
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Mus Mus

201991 rows × 13 columns

In [162]:
```
df4 = df.groupby(["type","country"])["title"].nunique().sort_values(ascending=False)
```

In [163]:
```
df4.rename(columns={"title":"count"},inplace=True)
```

In [164]:
```
df4
```

Out[164]:

	type	country	count
0	Movie	United States	3191
1	TV Show	United States	1329
2	Movie	India	962
3	Movie	United Kingdom	532
4	Movie	Canada	319
...	...	...	...
183	Movie	Paraguay	1
184	Movie	Panama	1
185	Movie	Palestine	1
186	Movie	Nicaragua	1
187	Movie	Slovakia	1

188 rows × 3 columns

Insight- Netflix has maximum number of content belongs United States

Analysis-10 Popular listed_in based on country on Netflix

In [165]:
```
df
```

Out[165]:

	show_id	type	title	director	cast	country	release_year	rating	duration	liste
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Liam Neeson	United States	2020	PG-13	90	Documenta
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	Internati TV Sh
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Dra
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Myste
1	s2	TV Show	Blood & Water	Martin Scorsese	Khosi Ngema	South Africa	2021	TV-MA	2	Internati TV Sh
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2015	TV-14	111	Internati Mc
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2015	TV-14	111	Mus Mus
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Dra
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Internati Mc
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Mus Mus

201991 rows × 13 columns

In [166]:
```
df.groupby(["listed_in","country"])["title"].nunique().sort_values(ascending=False)
```

Out[166]:
```
listed_in               country
Dramas                  United States    945
International Movies    India            864
Comedies                United States    774
Dramas                  India            662
Documentaries           United States    586
                                        ...
International TV Shows                     1
                        Austria            1
                        Azerbaijan         1
                        Croatia            1
Thrillers               West Germany       1
Name: title, Length: 1422, dtype: int64
```

Insight- Most popular Genre on Netflix is Dramas

Analysis-11 Average Duration of Tv show and Movies on Netflix

In [167]:
```
df
```

Out[167]:

	show_id	type	title	director	cast	country	release_year	rating	duration	liste
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Liam Neeson	United States	2020	PG-13	90	Documenta
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	Internati TV Sh
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Dra
1	s2	TV Show	Blood & Water	Martin Scorsese	Ama Qamata	South Africa	2021	TV-MA	2	TV Myste
1	s2	TV Show	Blood & Water	Martin Scorsese	Khosi Ngema	South Africa	2021	TV-MA	2	Internati TV Sh
...	...	...	...	...	...	...	...	...	...	
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2015	TV-14	111	Internati Mo
8806	s8807	Movie	Zubaan	Mozez Singh	Anita Shabdish	India	2015	TV-14	111	Mus Mus
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Dra
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Internati Mo
8806	s8807	Movie	Zubaan	Mozez Singh	Chittaranjan Tripathy	India	2015	TV-14	111	Mus Mus

201991 rows × 13 columns

In [168]:
```
df["duration"].value_counts()
```

Out[168]:
```
1      35035
2       9559
3       5088
94      4343
106     4040
       ...
16         4
196        4
20         4
18         4
nan        3
Name: duration, Length: 211, dtype: int64
```

In [169]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 201991 entries, 0 to 8806
Data columns (total 13 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   show_id        201991 non-null  object 
 1   type           201991 non-null  object 
 2   title          201991 non-null  object 
 3   director       201991 non-null  object 
 4   cast           201991 non-null  object 
 5   country        201991 non-null  object 
 6   release_year   201991 non-null  int64  
 7   rating         201991 non-null  object 
 8   duration       201991 non-null  object 
 9   listed_in      201991 non-null  object 
 10  month          201991 non-null  float64
 11  year           201991 non-null  float64
 12  week_day_name  201991 non-null  object 
dtypes: float64(2), int64(1), object(10)
memory usage: 21.6+ MB
```

In [170]:
```
df["duration"].fillna(df["duration"].mode()[0],inplace=True)
```

In [171]:
```
df["duration"] = df["duration"].map(lambda x:0 if x=="nan" else x)
```

In [172]:
```
df["duration"].value_counts()
```

Out[172]:
```
1      35035
2       9559
3       5088
94      4343
106     4040
       ...
16         4
196        4
20         4
18         4
0          3
Name: duration, Length: 211, dtype: int64
```

In [173]:
```
df["duration"]=pd.to_numeric(df["duration"])
```

In [174]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 201991 entries, 0 to 8806
Data columns (total 13 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   show_id        201991 non-null  object 
 1   type           201991 non-null  object 
 2   title          201991 non-null  object 
 3   director       201991 non-null  object 
 4   cast           201991 non-null  object 
 5   country        201991 non-null  object 
 6   release_year   201991 non-null  int64  
 7   rating         201991 non-null  object 
 8   duration       201991 non-null  int64  
 9   listed_in      201991 non-null  object 
 10  month          201991 non-null  float64
 11  year           201991 non-null  float64
 12  week_day_name  201991 non-null  object 
dtypes: float64(2), int64(2), object(9)
memory usage: 21.6+ MB
```

In [175]:
```
df.groupby(["type"])["duration"].mean()
```

Out[175]:
```
type
Movie      106.854254
TV Show      1.928101
Name: duration, dtype: float64
```

Insight- Average duration for Movies is approx. 107 min and for Tv Show is approx. 2 season.

Recomondations

1- Netflix has to be focus on TV Show as well as Movies because Tv show is also on trending

2-Netflix has to be focus on other rating content like TV-Y7, TV-Y, TV-G.

3-After analysing the data movies release per year is gretaer than movie added on Netflix so try to add more content on Netflix according to release per year

In [ ]:
```

```