Problem statement: In this case study we are going to analyze and explore which are the factors and how based on those factors Netflix can produce shows/movies and if shows or movies backed by data. We will decide and help the stakeholders re-strategize the business and also grow it.

Basic Metrics: The dataset has a total of 8807 shows and a total of 12 columns/attributes which we are going to analyze to help us understand and decide what shows/movies to produce more for the growth of business. The dataset has movies/shows for a total period of 96 years and since it is a large number this will be very helpful and we can be more precise with our analysis.

In [108]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [109]:
```
df=pd.read_csv('C:/DSML/netflix.csv')
```

In [110]:
```
#Basic Metrics
df.shape
#no. of rows/no. of titles
rows = df.shape[0]
rows
df["title"].nunique()
cols = df.shape[1]
cols
#Total period over which movies/shows are available
df["release_year"].max() - df["release_year"].min()
```

Out[110]:
```
96
```

In [111]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
show_id         8807 non-null object
type            8807 non-null object
title           8807 non-null object
director        6173 non-null object
cast            7982 non-null object
country         7976 non-null object
date_added      8797 non-null object
release_year    8807 non-null int64
rating          8803 non-null object
duration        8804 non-null object
listed_in       8807 non-null object
description     8807 non-null object
dtypes: int64(1), object(11)
memory usage: 825.7+ KB
```

missing values treatment

In [112]:
```
# all missing values for director will be made not known since mode cannot be the most appropriate value
# and this data has a wide range
(df["director"].isnull().sum()/df.shape[0])*100
```

Out[112]:
```
29.908027705234474
```

In [113]:
```
df['director']=df['director'].transform(lambda x: x.fillna(value='not known'))
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
show_id         8807 non-null object
type            8807 non-null object
title           8807 non-null object
director        8807 non-null object
cast            7982 non-null object
country         7976 non-null object
date_added      8797 non-null object
release_year    8807 non-null int64
rating          8803 non-null object
duration        8804 non-null object
listed_in       8807 non-null object
description     8807 non-null object
dtypes: int64(1), object(11)
memory usage: 825.7+ KB
```

In [114]:
```
# all missing values for cast will be made not known since mode cannot be the most appropriate value and this
# data has a wide range
(df["cast"].isnull().sum()/df.shape[0])*100
```

Out[114]:
```
9.367548540933349
```

In [115]:
```
df['cast']=df['cast'].transform(lambda x: x.fillna(value='not known'))
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
show_id         8807 non-null object
type            8807 non-null object
title           8807 non-null object
director        8807 non-null object
cast            8807 non-null object
country         7976 non-null object
date_added      8797 non-null object
release_year    8807 non-null int64
rating          8803 non-null object
duration        8804 non-null object
listed_in       8807 non-null object
description     8807 non-null object
dtypes: int64(1), object(11)
memory usage: 825.7+ KB
```

In [116]:
```
#missing values for country will be filled with the most occuring country value
(df["country"].isnull().sum()/df.shape[0])*100
```

Out[116]:
```
9.435676166685592
```

In [117]:
```
df['country']=df['country'].transform(lambda x: x.fillna(x.mode()[0]))
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
show_id         8807 non-null object
type            8807 non-null object
title           8807 non-null object
director        8807 non-null object
cast            8807 non-null object
country         8807 non-null object
date_added      8797 non-null object
release_year    8807 non-null int64
rating          8803 non-null object
duration        8804 non-null object
listed_in       8807 non-null object
description     8807 non-null object
dtypes: int64(1), object(11)
memory usage: 825.7+ KB
```

In [118]:
```
(df["date_added"].isnull().sum()/df.shape[0])*100
```

Out[118]:
```
0.11354604292040424
```

In [119]:
```
(df["rating"].isnull().sum()/df.shape[0])*100
```

Out[119]:
```
0.04541841716816169
```

In [120]:
```
(df["duration"].isnull().sum()/df.shape[0])*100
```

Out[120]:
```
0.034063812876121265
```

In [121]:
```
#remaining rows will no values will be dropped because the percentage of missing values to less than 2% of total rows
df = df.dropna(how='any')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8790 entries, 0 to 8806
Data columns (total 12 columns):
show_id         8790 non-null object
type            8790 non-null object
title           8790 non-null object
director        8790 non-null object
cast            8790 non-null object
country         8790 non-null object
date_added      8790 non-null object
release_year    8790 non-null int64
rating          8790 non-null object
duration        8790 non-null object
listed_in       8790 non-null object
description     8790 non-null object
dtypes: int64(1), object(11)
memory usage: 892.7+ KB
```

pre-processing of data to draw meaningful insights

In [122]:
```
#splitting cast based on title since each TV Show/Movie can have multiple actors/cast
constraint=df['cast'].apply(lambda x: str(x).split(', ')).tolist()
df_cast=pd.DataFrame(constraint,index=df['title'])
df_cast = df_cast.stack()
df_cast=pd.DataFrame(df_cast)
df_cast.reset_index(inplace=True)
df_cast=df_cast[['title',0]]
df_cast.columns = ['title','cast']
df_cast.head(10)
```

Out[122]:

	title	cast
0	Dick Johnson Is Dead	not known
1	Blood & Water	Ama Qamata
2	Blood & Water	Khosi Ngema
3	Blood & Water	Gail Mabalane
4	Blood & Water	Thabang Molaba
5	Blood & Water	Dillon Windvogel
6	Blood & Water	Natasha Thahane
7	Blood & Water	Arno Greeff
8	Blood & Water	Xolile Tshabalala
9	Blood & Water	Getmore Sithole

In [123]:
```
#splitting director based on title since each TV Show/Movie can have multiple actors/cast
constraint1=df['director'].apply(lambda x: str(x).split(', ')).tolist()
df_director=pd.DataFrame(constraint1,index=df['title'])
df_director = df_director.stack()
df_director=pd.DataFrame(df_director)
df_director.reset_index(inplace=True)
df_director=df_director[['title',0]]
df_director.columns = ['title','director']
df_director.head(10)
```

Out[123]:

	title	director
0	Dick Johnson Is Dead	Kirsten Johnson
1	Blood & Water	not known
2	Ganglands	Julien Leclercq
3	Jailbirds New Orleans	not known
4	Kota Factory	not known
5	Midnight Mass	Mike Flanagan
6	My Little Pony: A New Generation	Robert Cullen
7	My Little Pony: A New Generation	José Luis Ucha
8	Sankofa	Haile Gerima
9	The Great British Baking Show	Andy Devonshire

In [124]:
```
#splitting country based on title since each TV Show/Movie can be available in more than one country
constraint2=df['country'].apply(lambda x: str(x).split(', ')).tolist()
df_country=pd.DataFrame(constraint2,index=df['title'])
df_country = df_country.stack()
df_country=pd.DataFrame(df_country)
df_country.reset_index(inplace=True)
df_country=df_country[['title',0]]
df_country.columns = ['title','country']
df_country.head(10)
```

Out[124]:

	title	country
0	Dick Johnson Is Dead	United States
1	Blood & Water	South Africa
2	Ganglands	United States
3	Jailbirds New Orleans	United States
4	Kota Factory	India
5	Midnight Mass	United States
6	My Little Pony: A New Generation	United States
7	Sankofa	United States
8	Sankofa	Ghana
9	Sankofa	Burkina Faso

In [125]:
```
df = df.merge(df_cast,on='title')
df.drop('cast_x',axis=1,inplace=True)
df.rename(columns = {'cast_y':'cast'}, inplace = True)
df.head(10)
```

Out[125]:

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in	description	cast
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	not known
1	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Ama Qamata
2	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Khosi Ngema
3	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Gail Mabalane
4	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Thabang Molaba
5	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Dillon Windvogel
6	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Natasha Thahane
7	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Arno Greeff
8	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Xolile Tshabalala
9	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Getmore Sithole

In [126]:
```
df = df.merge(df_director,on='title')
df.head(10)
```

Out[126]:

	show_id	type	title	director_x	country	date_added	release_year	rating	duration	listed_in	description	cast	director_y
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	not known	Kirsten Johnson
1	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Ama Qamata	not known
2	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Khosi Ngema	not known
3	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Gail Mabalane	not known
4	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Thabang Molaba	not known
5	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Dillon Windvogel	not known
6	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Natasha Thahane	not known
7	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Arno Greeff	not known
8	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Xolile Tshabalala	not known

	show_id	type	title	director_x	country	date_added	release_year	rating	duration	listed_in	description	cast	director_y
9	s2	TV Show	Blood & Water	not known	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Getmore Sithole	not known

In [127]:
```
df.drop('director_x',axis=1,inplace=True)
df.rename(columns = {'director_y':'director'}, inplace = True)
df.head(10)
```

Out[127]:

	show_id	type	title	country	date_added	release_year	rating	duration	listed_in	description	cast	director
0	s1	Movie	Dick Johnson Is Dead	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	not known	Kirsten Johnson
1	s2	TV Show	Blood & Water	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Ama Qamata	not known
2	s2	TV Show	Blood & Water	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Khosi Ngema	not known
3	s2	TV Show	Blood & Water	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Gail Mabalane	not known
4	s2	TV Show	Blood & Water	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Thabang Molaba	not known
5	s2	TV Show	Blood & Water	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Dillon Windvogel	not known
6	s2	TV Show	Blood & Water	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Natasha Thahane	not known
7	s2	TV Show	Blood & Water	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Arno Greeff	not known
8	s2	TV Show	Blood & Water	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Xolile Tshabalala	not known
9	s2	TV Show	Blood & Water	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Getmore Sithole	not known

In [128]:
```
df = df.merge(df_country,on='title')
df.drop('country_x',axis=1,inplace=True)
df.rename(columns = {'country_y':'country'}, inplace = True)
df.head(10)
```

Out[128]:

	show_id	type	title	date_added	release_year	rating	duration	listed_in	description	cast	director	country
0	s1	Movie	Dick Johnson Is Dead	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	not known	Kirsten Johnson	United States
1	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Ama Qamata	not known	South Africa
2	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Khosi Ngema	not known	South Africa
3	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Gail Mabalane	not known	South Africa
4	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Thabang Molaba	not known	South Africa
5	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Dillon Windvogel	not known	South Africa
6	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Natasha Thahane	not known	South Africa
7	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Arno Greeff	not known	South Africa
8	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Xolile Tshabalala	not known	South Africa
9	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	Getmore Sithole	not known	South Africa

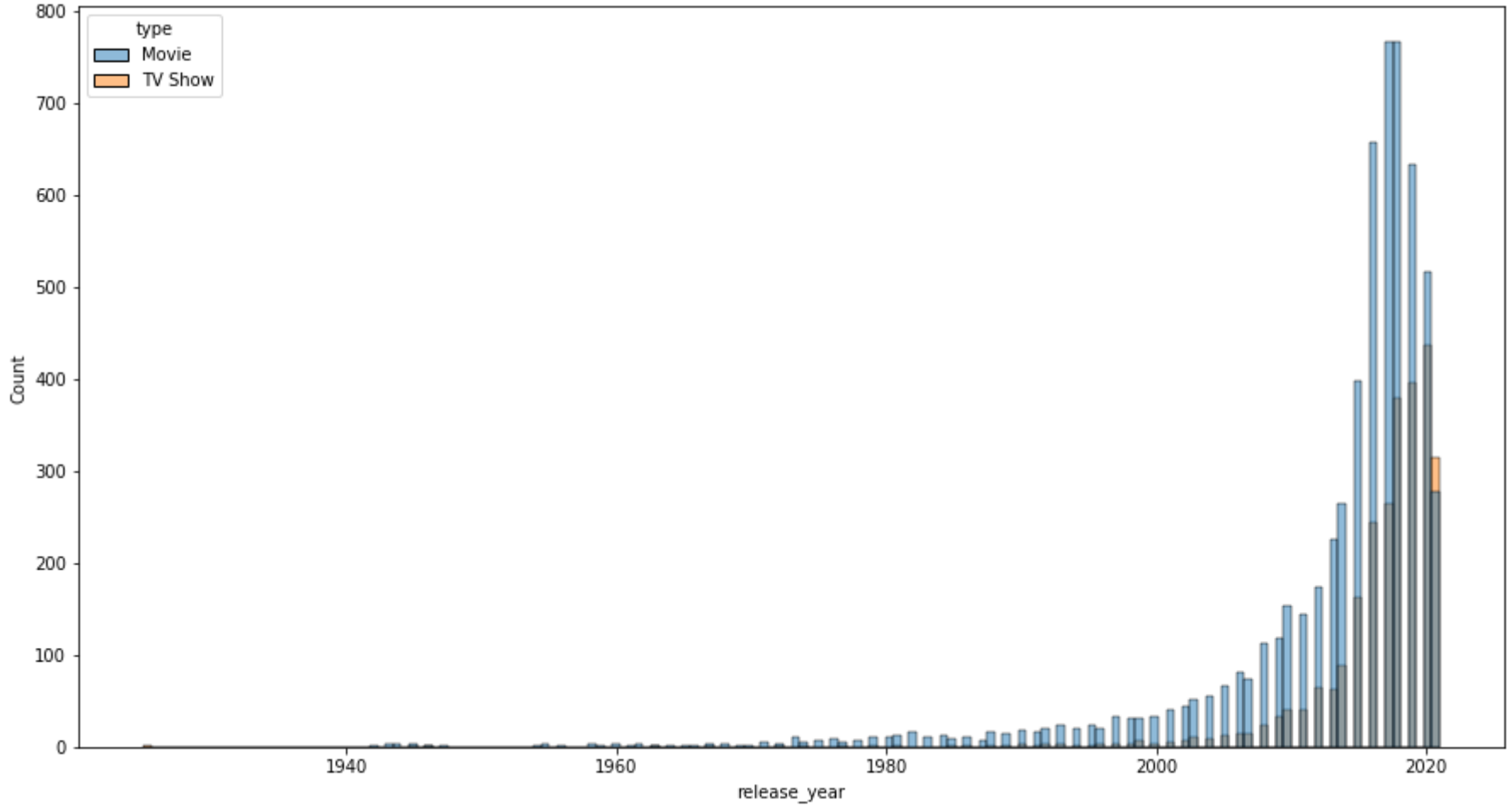
In [129]:
```
#TV shows vs Movies
df.groupby('type')['title'].nunique()
```

Out[129]:
```
type
Movie      6126
TV Show    2664
Name: title, dtype: int64
```

visual analysis

In [130]:
```
import seaborn as sns
import matplotlib.pyplot as plt
df_old = pd.read_csv('C:/DSML/netflix.csv')
plt.figure(figsize=(15,8))
sns.histplot(x='release_year',hue='type',data=df_old)
```

Out[130]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x2b2d2e7fef0>
```



By looking at the above plot we can infer jthe below points

1.The number of movies that got released increased exponentially after the year 2000. But there has been a drastic decrease after the year 2017-18

2. As the no. of movies made started to decrease, we can also observe that there has been a steady increase in the no. of TV shows made(though there is a slight drop in no. of TV shows made in 2022, this observation will still hold valid). Hence we can also safely say by looking at the graph that the no of TV shows made is inversely varying with respect to the no. of Movies made since the year 2017-2018

In [131]:
```
df_old = pd.read_csv('C:/DSML/netflix.csv')
df_old['date_add']=pd.to_datetime(df_old['date_added'], errors='coerce')
df_old.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 13 columns):
show_id         8807 non-null object
type            8807 non-null object
title           8807 non-null object
director        6173 non-null object
cast            7982 non-null object
country         7976 non-null object
date_added      8797 non-null object
release_year    8807 non-null int64
rating          8803 non-null object
duration        8804 non-null object
listed_in       8807 non-null object
description     8807 non-null object
date_add        8797 non-null datetime64[ns]
dtypes: datetime64[ns](1), int64(1), object(11)
memory usage: 894.5+ KB
```

In [132]:
```
df_old.head()
```

Out[132]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	date_add
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, filmm...	2021-09-25
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	After crossing paths at a party, a Cape Town t...	2021-09-24
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 protect his family from a powerful drug lor...	2021-09-24
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 go down amo...	2021-09-24
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 to train I...	2021-09-24

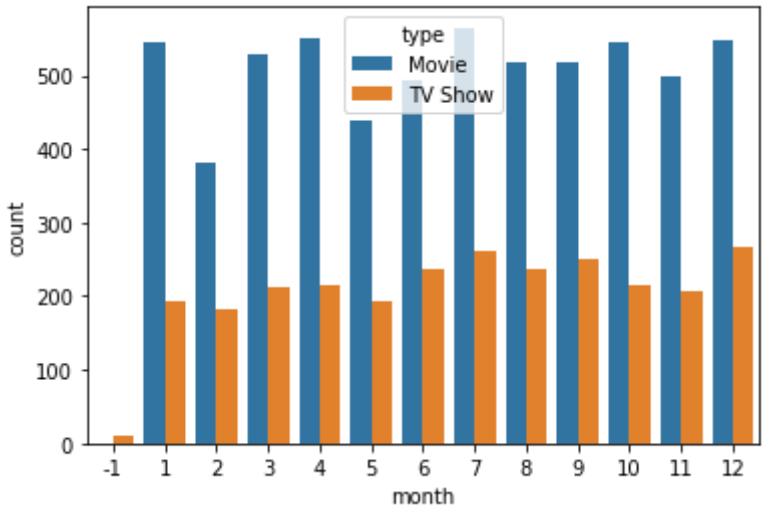
In [133]:
```
df_old['month']=df_old['date_add'].dt.month.fillna(-1)
```

In [134]:
```
df_old['month']=df_old['month'].astype('int64')
df_old
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	date_add
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, filmm...	2021-09-25
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	After crossing paths at a party, a Cape Town t...	2021-09-24
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 protect his family from a powerful drug lor...	2021-09-24
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 go	2021-09-24

In [135]:
```
#no of TV Shows/Movies
sns.countplot(x='month',hue='type', data=df_old)
```

Out[135]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x2b2d237e588>
```



By looking at the above plot we can infer the below factors

1. The no. of Movies released is always more than the no. of TV shows released every month
2. The best time to release a TV show would be from june-september and again in december. Although any month does not make a huge difference or has a big impact

In [136]:
```
constraint3=df['listed_in'].apply(lambda x: str(x).split(', ')).tolist()
df_genre=pd.DataFrame(constraint3,index=df['title'])
df_genre = df_genre.stack()
df_genre=pd.DataFrame(df_genre)
df_genre.reset_index(inplace=True)
df_genre=df_genre[['title',0]]
df_genre.columns = ['title','listed_in']
df_genre.head(10)
```

Out[136]:

	title	listed_in
0	Dick Johnson Is Dead	Documentaries
1	Blood & Water	International TV Shows
2	Blood & Water	TV Dramas
3	Blood & Water	TV Mysteries
4	Blood & Water	International TV Shows
5	Blood & Water	TV Dramas
6	Blood & Water	TV Mysteries
7	Blood & Water	International TV Shows
8	Blood & Water	TV Dramas
9	Blood & Water	TV Mysteries

In [137]:
```
df = df.merge(df_genre,on='title')
df.drop('listed_in_x',axis=1,inplace=True)
df.rename(columns = {'listed_in_y':'listed_in'}, inplace = True)
df
```

Out[137]:

	show_id	type	title	date_added	release_year	rating	duration	description	cast	director	country	listed_in
0	s1	Movie	Dick Johnson Is Dead	September 25, 2021	2020	PG-13	90 min	As her father nears the end of his life, filmm...	not known	Kirsten Johnson	United States	Documentaries
1	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Ama Qamata	not known	South Africa	International TV Shows
2	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Ama Qamata	not known	South Africa	TV Dramas
3	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Ama Qamata	not known	South Africa	TV Mysteries
4	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Ama Qamata	not known	South Africa	International TV Shows
		TV	Blood &	September		TV-	2	After crossing	Ama	not	South	

In [138]:
```
# Analysis of actors(cast)/director of different types of genres of shows/movies
#Top actors on Netflix
df.groupby('cast')['title'].nunique().sort_values(ascending=False).head(11)
```

Out[138]:
```
cast
not known           825
Anupam Kher          43
Shah Rukh Khan       35
Julie Tejwani        33
Takahiro Sakurai     32
Naseeruddin Shah     32
Rupa Bhimani         31
Akshay Kumar         30
Om Puri              30
Yuki Kaji            29
Amitabh Bachchan     28
Name: title, dtype: int64
```

In [139]:
```
#Top directors on Netflix
df.groupby('director')['title'].nunique().sort_values(ascending=False).head(11)
```

Out[139]:
```
director
not known              2621
Rajiv Chilaka            22
Jan Suter                21
Raúl Campos              19
Suhas Kadav              16
Marcus Raboy             16
Jay Karas                15
Cathy Garcia-Molina      13
Jay Chapman              12
Youssef Chahine          12
Martin Scorsese          12
Name: title, dtype: int64
```

In [140]:
```
# popular actor-director combination
df.groupby(['cast','director'])['title'].nunique().sort_values(ascending=False).head(10)
```

Out[140]:
```
cast              director
not known         not known        352
Takahiro Sakurai  not known         24
Rajesh Kava       Rajiv Chilaka     19
Julie Tejwani     Rajiv Chilaka     19
Yuki Kaji         not known         18
Rupa Bhimani      Rajiv Chilaka     18
Jigna Bhardwaj    Rajiv Chilaka     18
Yuichi Nakamura   not known         16
Daisuke Ono       not known         16
Junichi Suwabe    not known         16
Name: title, dtype: int64
```

In [141]:
```
#Top genres on Netflix
df.groupby(['listed_in','type'])['title'].nunique().sort_values(ascending=False).head(10)
```

Out[141]:
```
listed_in                type
International Movies     Movie      2752
Dramas                   Movie      2426
Comedies                 Movie      1674
International TV Shows   TV Show    1349
Documentaries            Movie       869
Action & Adventure       Movie       859
TV Dramas                TV Show     762
Independent Movies       Movie       756
Children & Family Movies Movie       641
Romantic Movies          Movie       616
Name: title, dtype: int64
```

In [142]:
```
# top genre of movies/shows in different countries
df.groupby(['listed_in','type'])[['title','country']].nunique().sort_values(by=['title','country'],ascending=False).head
```

Out[142]:

listed_in	type	title	country
International Movies	Movie	2752	104
Dramas	Movie	2426	96
Comedies	Movie	1674	66
International TV Shows	TV Show	1349	62
Documentaries	Movie	869	75
Action & Adventure	Movie	859	59
TV Dramas	TV Show	762	60
Independent Movies	Movie	756	67
Children & Family Movies	Movie	641	48
Romantic Movies	Movie	616	53

In [143]:
```
# most popular actor for International Movies
df_internl_movies = df_cast.merge(df_genre,on='title')
df_internl_movies = df_internl_movies.loc[df_internl_movies["listed_in"] == 'International Movies']
df_internl_movies = df_internl_movies.drop_duplicates(keep='first')
#taking the second actor from the list since the top most cast is not known
df_internl_movies['cast'].value_counts().index.tolist()[1]
```

Out[143]: 'Anupam Kher'

In [144]:
```
# most popular actor for Dramas
df_dramas = df_cast.merge(df_genre,on='title')
df_dramas = df_dramas.loc[df_dramas["listed_in"] == 'Dramas']
df_dramas = df_dramas.drop_duplicates(keep='first')
#taking the second actor from the list since the top most cast is not known
df_internl_movies['cast'].value_counts().index.tolist()[1]
```

Out[144]: 'Anupam Kher'

In [145]:
```
# most popular actor for comedy
df_comedy = df_cast.merge(df_genre,on='title')
df_comedy = df_comedy.loc[df_comedy["listed_in"] == 'Comedies']
df_comedy = df_comedy.drop_duplicates(keep='first')
df_comedy['cast'].value_counts().index.tolist()[0]
```

Out[145]: 'Anupam Kher'

In [146]:
```
# most popular actor for International TV shows
df_internl_TV_shows = df_cast.merge(df_genre,on='title')
df_internl_TV_shows = df_internl_TV_shows.loc[df_internl_TV_shows["listed_in"] == 'International TV Shows']
df_internl_TV_shows = df_internl_TV_shows.drop_duplicates(keep='first')
#taking the second actor from the list since the top most cast is not known
df_internl_TV_shows['cast'].value_counts().index.tolist()[1]
```

Out[146]: 'Takahiro Sakurai'

In [147]:
```
# most popular actor for Documentaries
df_docum = df_cast.merge(df_genre,on='title')
df_docum = df_docum.loc[df_docum["listed_in"] == 'Documentaries']
df_docum = df_docum.drop_duplicates(keep='first')
#taking the second actor from the list since the top most cast is not known
df_docum['cast'].value_counts().index.tolist()[1]
```

Out[147]: 'Samuel West'

In [148]:
```
#actor with most screen time in Movies
df_dur = df_old.loc[df_old['type']=='Movie']
constraint4=df_dur['duration'].apply(lambda x: str(x).split(' ')).tolist()
df_duration=pd.DataFrame(constraint4,index=df_dur['title'])
df_duration.drop(df_duration.iloc[:, 1:], inplace = True, axis = 1)
df_duration = df_duration.stack()
df_duration=pd.DataFrame(df_duration)
df_duration.reset_index(inplace=True)
df_duration=df_duration[['title',0]]
df_duration.columns = ['title','duration']
df_most_st = df_cast.merge(df_duration,on='title')
df_most_st['duration']=df_most_st['duration'].astype('int64')
df_most_st.groupby('cast')['duration'].sum().sort_values(ascending=False).head(2)
```

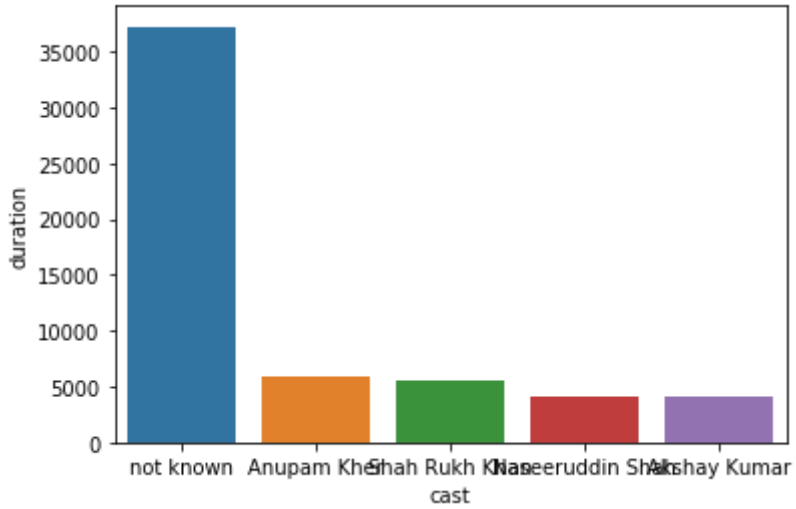
Out[148]:
```
cast
not known      37242
Anupam Kher     5929
Name: duration, dtype: int64
```

In [149]:
```
# visual represantation of actors screen time in descending order(top 5)
```

In [150]:
```
df_most_st_top = df_most_st.groupby('cast')['duration'].sum().sort_values(ascending=False).head(5)
df_most_st_top = pd.DataFrame(df_most_st_top)
df_most_st_top.reset_index(inplace=True)
```

In [151]:
```
sns.barplot(data=df_most_st_top,
            x="cast",
            y="duration")
```

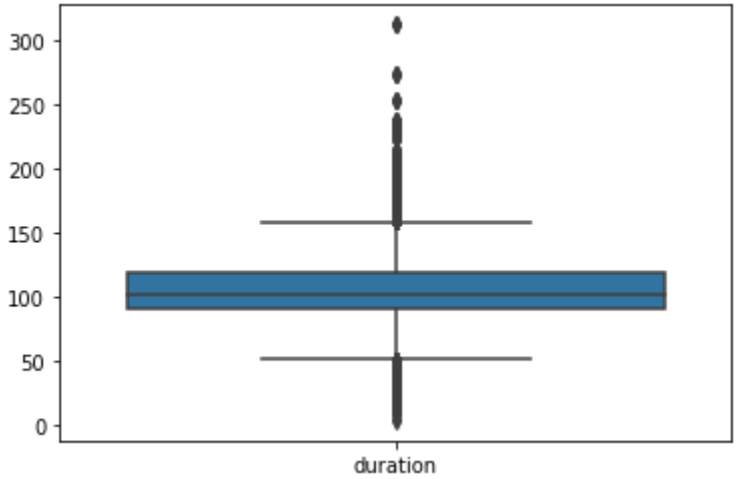
Out[151]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x2b28204b4e0>
```



The above plot depicts the total screen time of movies of top 5 actors on Netflix. Although we have a lot of movies with no cast details given, we can conclude that 'Anupam Kher' and 'Shahrukh Khan' share almost the same amount of screen time followed by 'Naseeruddin Shah' and 'Akshay kumar'

In [152]:
```
sns.boxplot(data=df_most_st)
```

Out[152]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x2b2cf266f60>
```



The above plot depicts the total screen time duration for movies of each of the actor from the data set. By looking at the graph we can say that the most screen time duration lies between around 90 and 120 minutes. Although there are a lot of outliers this graph can still be used to understand that the average screentime duration does not vary a lot for all the actors

In [ ]:

In [ ]: