

```
In [245... import gdown
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statistics
```

```
In [246... ! gdown https://d2beiqrkhq929f0.cloudfront.net/public_assets/assets/000/000/940/original/netflix.csv

Downloading...
From: https://d2beiqrkhq929f0.cloudfront.net/public_assets/assets/000/000/940/original/netflix.csv
To: /content/netflix.csv
100% 3.40M/3.40M [00:00<00:00, 26.7MB/s]
```

```
In [247... data = pd.read_csv('netflix.csv')
data
```

Out[247]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
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...
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...
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...
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...
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 l...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...

8807 rows x 12 columns

- 1)Netflix have two type of show ( movies and TV show)
- 2)There are total 8807 show (movies and TV show) each with unique show\_id and title
- 3)Here cast and director are in concat form exact count is still unknown
- 4)This is 74 years of data
- 5)Netflix is active approx(because country have null values) 74 country

```
In [295... data.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 [296... data.keys()

Out[296]: Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
      'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object')
```

```
In [297... data.notnull().sum(axis=0) #Identifying the null values

Out[297]: show_id      8807
type          8807
title         8807
director      6173
cast          7982
country       7976
date_added    8797
release_year  8807
rating        8803
duration      8804
listed_in     8807
description   8807
dtype: int64
```

## Data Filtration

Here we are seperating Cast column and filtering it seperately.

```
In [248... Cast=data[["show_id","cast"]]
Cast["cast"]=Cast["cast"].str.split(",") #Splitting with comma
Cast=Cast.explode("cast")
Cast["cast"]=Cast["cast"].str.lower().str.strip()
Cast.shape
Cast["cast"].value_counts()

<ipython-input-248-5c962278a0d6>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
Cast["cast"]=Cast["cast"].str.split(",") #Splitting with comma

Out[248]: anupam kher      43
shah rukh khan   35
julie tejwani    33
takahiro sakurai 32
naseeruddin shah 32
..
melanie straub   1
gabriela maria schmeide 1
helena zengel    1
daniel valenzuela 1
chittaranjan tripathy 1
Name: cast, Length: 36428, dtype: int64
```

Here we are seperating Country column and filtering it seperately.

In [249..

```
country=data[["show_id","country"]]
country["country"]=country["country"].str.split(",") #Splitting with comma
country=country.explode("country")
country["country"]=country["country"].str.lower().str.strip()
country["country"].value_counts()
```

```
<ipython-input-249-06beb238bcd>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
country["country"]=country["country"].str.split(",") #Splitting with comma
```

Out[249]:

```
united states    3690
india            1046
united kingdom   806
canada           445
france           393
...
ecuador          1
armenia          1
mongolia         1
bahamas          1
montenegro       1
Name: country, Length: 123, dtype: int64
```

Here we are separating Listed\_in column and filtering it separately.

In [250..

```
listed_in=data[["show_id","listed_in"]]
#Listed_in
listed_in["listed_in"]=listed_in["listed_in"].str.split(",") #Splitting with comma
#nListed_in
listed_in["listed_in"]=listed_in["listed_in"].explode("listed_in")
listed_in["listed_in"]=listed_in["listed_in"].str.lower().str.strip()
listed_in["listed_in"].value_counts()
```

```
<ipython-input-250-13cbcd6b7379>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
listed_in["listed_in"]=listed_in["listed_in"].str.split(",") #Splitting with comma
```

```
<ipython-input-250-13cbcd6b7379>:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
listed_in["listed_in"]=listed_in["listed_in"].explode("listed_in")
```

```
<ipython-input-250-13cbcd6b7379>:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
listed_in["listed_in"]=listed_in["listed_in"].str.lower().str.strip()
```

```
Out[250]: international movies      1231
dramas                          1029
comedies                        761
international tv shows          725
tv dramas                       409
action & adventure              346
tv comedies                     322
children & family movies        313
romantic movies                 303
documentaries                   278
independent movies              259
thrillers                       255
crime tv shows                  250
kids' tv                        248
romantic tv shows               189
docuseries                      187
music & musicals                 159
reality tv                      157
horror movies                   125
anime series                    120
tv action & adventure            112
stand-up comedy                 111
spanish-language tv shows       108
british tv shows                 90
korean tv shows                 78
sports movies                   75
sci-fi & fantasy                 73
tv mysteries                     59
tv sci-fi & fantasy              56
teen tv shows                   53
tv horror                       47
lgbtq movies                    46
anime features                  46
science & nature tv              37
tv thrillers                    34
classic movies                  30
stand-up comedy & talk shows     21
faith & spirituality             18
cult movies                     18
movies                          12
classic & cult tv                 9
tv shows                        8
Name: listed_in, dtype: int64
```

After separately filtering columns we are joining all the datasets.

```
In [251... # Merging All the Datasets using inner join
df1=Cast.merge(country, how="inner",on="show_id")
df1.shape
df2=df1.merge(listed_in, how="inner",on="show_id")
df2
```

```
Out[251]:
```

	show_id	cast	country	listed_in
0	s1	NaN	united states	documentaries
1	s2	ama qamata	south africa	international tv shows
2	s2	khosi ngema	south africa	international tv shows
3	s2	gail mabalane	south africa	international tv shows
4	s2	thabang molaba	south africa	international tv shows
...	...	...	...	...
81736	s8807	manish chaudhary	india	international tv shows
81737	s8807	meghna malik	india	international tv shows
81738	s8807	malkeet rauni	india	international tv shows
81739	s8807	anita shabdish	india	international tv shows
81740	s8807	chittaranjan tripathy	india	international tv shows

81741 rows × 4 columns

```
In [252... #checking duplicate values in newly created dataset
df2[df2.duplicated()]
```

```
Out[252]:
```

	show_id	cast	country	listed_in
16460	s1632	micah hauptman	united states	comedies
54460	s6014	adrianna biedrzyńska	denmark	international movies
54461	s6014	adrianna biedrzyńska	france	international movies
54462	s6014	adrianna biedrzyńska	poland	international movies

```
In [253... #Handeling Duplucate values
df2=df2.drop_duplicates()
df2
```

```
Out[253]:
```

	show_id	cast	country	listed_in
0	s1	NaN	united states	documentaries
1	s2	ama qamata	south africa	international tv shows
2	s2	khosi ngema	south africa	international tv shows
3	s2	gail mabalane	south africa	international tv shows
4	s2	thabang molaba	south africa	international tv shows
...	...	...	...	...
81736	s8807	manish chaudhary	india	international tv shows
81737	s8807	meghna malik	india	international tv shows
81738	s8807	malkeet rauni	india	international tv shows
81739	s8807	anita shabdish	india	international tv shows
81740	s8807	chittaranjan tripathy	india	international tv shows

81737 rows × 4 columns

```
In [254... df=data.drop(['cast','country','listed_in','description'],axis=1) #removing cast, country, listed_in column from main datafr
df
```

```
Out[254]:
```

	show_id	type	title	director	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	September 25, 2021	2020	PG-13	90 min
1	s2	TV Show	Blood & Water	NaN	September 24, 2021	2021	TV-MA	2 Seasons
2	s3	TV Show	Ganglands	Julien Leclercq	September 24, 2021	2021	TV-MA	1 Season
3	s4	TV Show	Jailbirds New Orleans	NaN	September 24, 2021	2021	TV-MA	1 Season
4	s5	TV Show	Kota Factory	NaN	September 24, 2021	2021	TV-MA	2 Seasons
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	November 20, 2019	2007	R	158 min
8803	s8804	TV Show	Zombie Dumb	NaN	July 1, 2019	2018	TV-Y7	2 Seasons
8804	s8805	Movie	Zombieland	Ruben Fleischer	November 1, 2019	2009	R	88 min
8805	s8806	Movie	Zoom	Peter Hewitt	January 11, 2020	2006	PG	88 min
8806	s8807	Movie	Zubaan	Mozes Singh	March 2, 2019	2015	TV-14	111 min

8807 rows × 8 columns

After all the filtration we are combining all the small datasets and creating one ideal dataset for through out reference.

```
In [255... IDF=df.merge(df2,how="inner",on="show_id") #Merging newly created dataFrame with ideal dataFrame
IDF
```

Out[255]:

	show_id	type	title	director	date_added	release_year	rating	duration	cast	country	listed_in	
	0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	September 25, 2021	2020	PG-13	90 min	NaN	united states	documentaries
	1	s2	TV Show	Blood & Water	NaN	September 24, 2021	2021	TV- MA	2 Seasons	ama qamata	south africa	international tv shows
	2	s2	TV Show	Blood & Water	NaN	September 24, 2021	2021	TV- MA	2 Seasons	khosi ngema	south africa	international tv shows
	3	s2	TV Show	Blood & Water	NaN	September 24, 2021	2021	TV- MA	2 Seasons	gail mabalane	south africa	international tv shows
	4	s2	TV Show	Blood & Water	NaN	September 24, 2021	2021	TV- MA	2 Seasons	thabang molaba	south africa	international tv shows
	...	...	...	...	...	...	...	...	...	...	...	...
	81732	s8807	Movie	Zubaan	Mozez Singh	March 2, 2019	2015	TV-14	111 min	manish chaudhary	india	international tv shows
	81733	s8807	Movie	Zubaan	Mozez Singh	March 2, 2019	2015	TV-14	111 min	meghna malik	india	international tv shows
	81734	s8807	Movie	Zubaan	Mozez Singh	March 2, 2019	2015	TV-14	111 min	malkeet rauni	india	international tv shows
	81735	s8807	Movie	Zubaan	Mozez Singh	March 2, 2019	2015	TV-14	111 min	anita shabdish	india	international tv shows
	81736	s8807	Movie	Zubaan	Mozez Singh	March 2, 2019	2015	TV-14	111 min	chittaranjan tripathy	india	international tv shows

81737 rows × 11 columns

In [256...

```
IDF.isna().sum() #Checking Null values in dataset
```

Out[256]:

```
show_id      0
type          0
title         0
director    21937
date_added    69
release_year   0
rating        38
duration        3
cast         1006
country       5053
listed_in     0
dtype: int64
```

In [257...

```
#Handling null values in Director column
mode_value=IDF["director"].mode()[0]
IDF["director"].fillna(mode_value, inplace=True)

#Handling null values in Date_added column
mode_value=IDF["date_added"].mode()[0]
IDF["date_added"].fillna(mode_value, inplace=True)

#Handling null values in rating column
mode_value=IDF["rating"].mode()[0]
IDF["rating"].fillna(mode_value, inplace=True)

#Handling null values in duration column
mode_value=IDF["duration"].mode()[0]
IDF["duration"].fillna(mode_value, inplace=True)

#Handling null values in cast column
mode_value=IDF["cast"].mode()[0]
IDF["cast"].fillna(mode_value, inplace=True)

#Handling null values in country column
mode_value=IDF["country"].mode()[0]
IDF["country"].fillna(mode_value, inplace=True)
```

In [258...

```
IDF.isin([""]).sum() #Checking empty values
```

Out[258]:

```
show_id      0
type          0
title         0
director      0
date_added    0
release_year   0
rating         0
duration       0
cast           0
country       45
listed_in     0
dtype: int64
```

```
In [259... IDF["country"]=IDF["country"].replace("",IDF["country"].mode()[0])
IDF.isin([""]).sum()
```

```
Out[259]: show_id      0
type        0
title       0
director    0
date_added  0
release_year 0
rating      0
duration    0
cast        0
country     0
listed_in   0
dtype: int64
```

```
In [260... IDF.isnull().sum() #Final check for null values
```

```
Out[260]: show_id      0
type        0
title       0
director    0
date_added  0
release_year 0
rating      0
duration    0
cast        0
country     0
listed_in   0
dtype: int64
```

```
In [261... IDF.head()
```

```
Out[261]:
```

	show_id	type	title	director	date_added	release_year	rating	duration	cast	country	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	September 25, 2021	2020	PG-13	90 min	anupam kher	united states	documentaries
1	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	ama qamata	south africa	international tv shows
2	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	khosi ngema	south africa	international tv shows
3	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	gail mabalane	south africa	international tv shows
4	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	thabang molaba	south africa	international tv shows

## #Finding the counts of each categorical variable both using graphical and non- graphical analysis.

```
In [262... #For Non-graphical Analysis:
shows=IDF[["show_id","type","release_year","date_added","country"]].drop_duplicates(keep="first").reset_index(drop=True)
shows["type"].value_counts()
```

```
Out[262]: Movie      7818
TV Show    3031
Name: type, dtype: int64
```

```
In [263... movies=shows[shows["type"]=="Movie"].reset_index(drop=True)
tv_shows=shows[shows["type"]=="TV Show"].reset_index(drop=True)
```

```
In [264... tv_shows.head()
```

```
Out[264]:
```

	show_id	type	release_year	date_added	country
0	s2	TV Show	2021	September 24, 2021	south africa
1	s3	TV Show	2021	September 24, 2021	united states
2	s4	TV Show	2021	September 24, 2021	united states
3	s5	TV Show	2021	September 24, 2021	india
4	s6	TV Show	2021	September 24, 2021	united states

```
In [265... movies.head()
```

```
Out[265]:
```

	show_id	type	release_year	date_added	country
0	s1	Movie	2020	September 25, 2021	united states
1	s7	Movie	2021	September 24, 2021	united states
2	s8	Movie	1993	September 24, 2021	united states
3	s8	Movie	1993	September 24, 2021	ghana
4	s8	Movie	1993	September 24, 2021	burkina faso

```
In [266... #For Non-graphical Analysis:
movie_counts=pd.DataFrame(movies.groupby("release_year").size().iloc [-30:-1].reset_index().rename({0:"Count"},axis=1))
movie_counts
```

```
Out[266]:
```

	release_year	Count
0	1992	28
1	1993	40
2	1994	23
3	1995	24
4	1996	21
5	1997	42
6	1998	38
7	1999	42
8	2000	46
9	2001	50
10	2002	66
11	2003	77
12	2004	72
13	2005	100
14	2006	109
15	2007	96
16	2008	140
17	2009	168
18	2010	215
19	2011	194
20	2012	231
21	2013	296
22	2014	385
23	2015	544
24	2016	878
25	2017	974
26	2018	964
27	2019	763
28	2020	565

```
In [267... #For Non-graphical Analysis:
tvshow_counts=pd.DataFrame(tv_shows.groupby("release_year").size().iloc [-30:-1].reset_index().rename({0:"Count"},axis=1))
tvshow_counts
```



Out[267]:

	release_year	Count
0	1992	3
1	1993	6
2	1994	2
3	1995	2
4	1996	5
5	1997	10
6	1998	6
7	1999	9
8	2000	6
9	2001	5
10	2002	10
11	2003	12
12	2004	9
13	2005	15
14	2006	17
15	2007	18
16	2008	25
17	2009	37
18	2010	47
19	2011	53
20	2012	71
21	2013	70
22	2014	96
23	2015	184
24	2016	277
25	2017	302
26	2018	427
27	2019	445
28	2020	490

In [268...

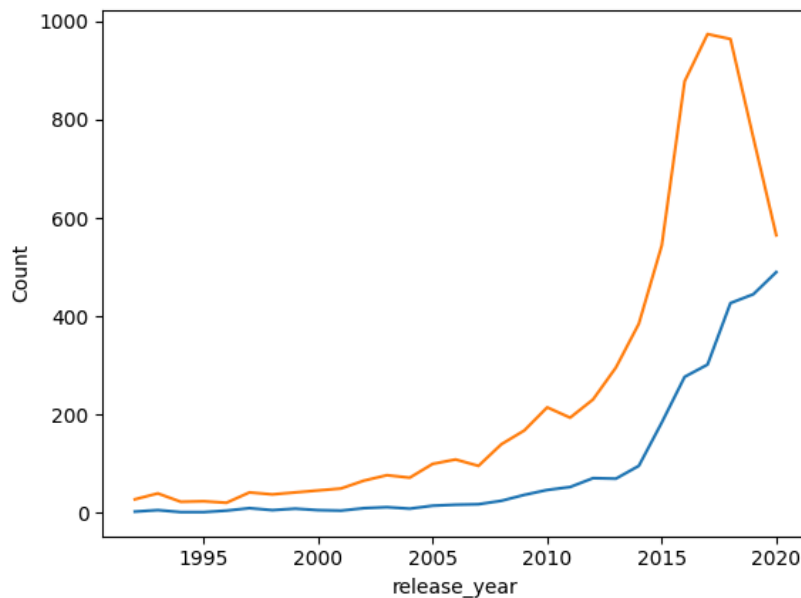
shows.head()

Out[268]:

	show_id	type	release_year	date_added	country
0	s1	Movie	2020	September 25, 2021	united states
1	s2	TV Show	2021	September 24, 2021	south africa
2	s3	TV Show	2021	September 24, 2021	united states
3	s4	TV Show	2021	September 24, 2021	united states
4	s5	TV Show	2021	September 24, 2021	india

In [269...

```
#graphical Analysis:
sns.lineplot(data=tvshow_counts,x="release_year",y="Count")
sns.lineplot(data=movie_counts,x="release_year",y="Count")
plt.show()
```



\*Number of Tv shows and Movies has been increased over the time.

\*After 2010 we can see the sudden increase in both Tv shows and Movies.

\*After 2015 the number of movies release started falling and number of Tv Shows release keep growing.

## Comparison in between TV Shows and Movies.

```
In [270... shows["type"].value_counts()
```

```
#So untill now 6131 Movies and 2676 TV Shows were released on Netflix
```

```
Out[270]: Movie      7818
TV Show    3031
Name: type, dtype: int64
```

```
In [271... #Top 10 country where most movies is release
```

```
temp_country=IDF[["show_id","type","country"]].drop_duplicates(keep="first").reset_index(drop=True)
temp_country[temp_country["type"]=="Movie"]["country"].value_counts()[:10]
```

```
Out[271]: united states    3197
india                962
united kingdom       534
canada               319
france               303
germany              182
spain                171
japan                119
china                114
mexico              111
Name: country, dtype: int64
```

```
In [272... #Top 10 country where most TV Shows is release
```

```
temp_country[temp_country["type"]=="TV Show"]["country"].value_counts()[:10]
```

```
Out[272]: united states    1330
united kingdom    272
japan             199
south korea       170
canada            126
france            90
india             84
taiwan            70
australia         66
spain             61
Name: country, dtype: int64
```

## The best time to launch a TV show.

```
In [273... IDF.head()
```

Out[273]:	show_id	type	title	director	date_added	release_year	rating	duration	cast	country	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	September 25, 2021	2020	PG-13	90 min	anupam kher	united states	documentaries
1	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	ama qamata	south africa	international tv shows
2	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	khosi ngema	south africa	international tv shows
3	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	gail mabalane	south africa	international tv shows
4	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	thabang molaba	south africa	international tv shows

```
In [274... # Data of Top 5 Years
top_years=tv_shows["release_year"].value_counts().index[:5]
top_tv_shows=IDF[IDF["release_year"].isin(top_years) & (IDF["type"]=="TV Show")][["show_id","release_year","date_added","country"]]
top_tv_shows.drop_duplicates(keep="first",inplace=True)
top_tv_shows.reset_index(drop=True,inplace=True)
top_tv_shows["date_added"]=pd.to_datetime(top_tv_shows["date_added"])
top_tv_shows.head()
```

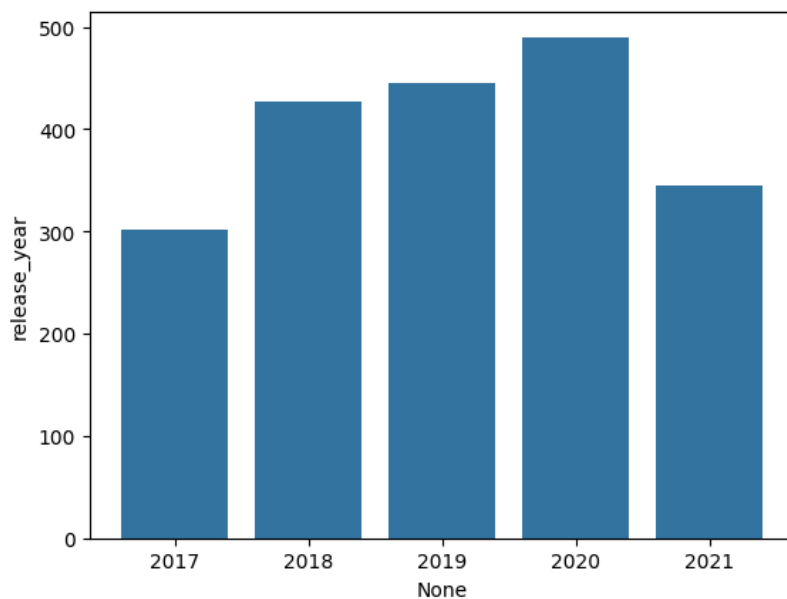
Out[274]:	show_id	release_year	date_added	country
0	s2	2021	2021-09-24	south africa
1	s3	2021	2021-09-24	united states
2	s4	2021	2021-09-24	united states
3	s5	2021	2021-09-24	india
4	s6	2021	2021-09-24	united states

```
In [275... # As we are observing most number of tv Show released on 2020
year=tv_shows["release_year"].value_counts()[:5]
year
```

```
Out[275]: 2020    490
2019    445
2018    427
2021    345
2017    302
Name: release_year, dtype: int64
```

```
In [276... sns.barplot(x=year.index,y=year)
plt.plot()
```

Out[276]: []

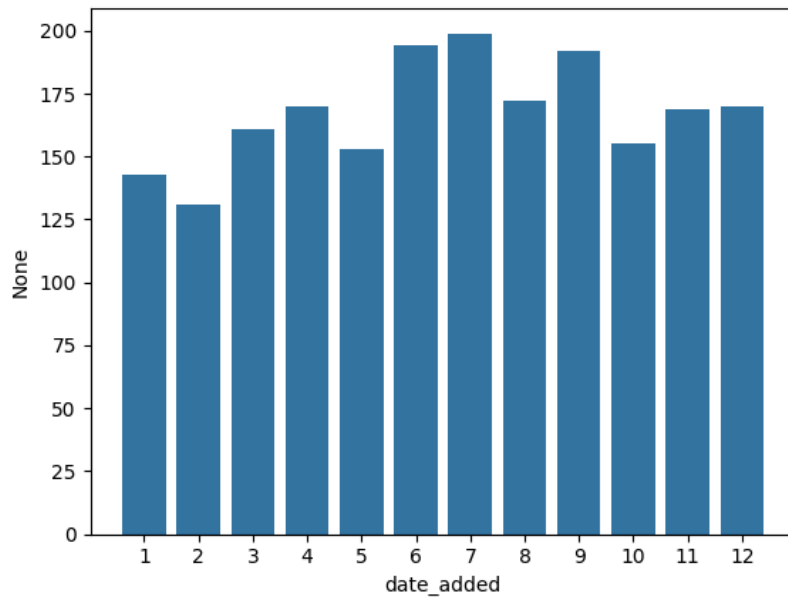


```
In [277... # 6,5,9 th month are have most number of tv_show
month=top_tv_shows.groupby(top_tv_shows["date_added"].dt.month).size()
month
```

```
Out[277]: date_added
1      143
2      131
3      161
4      170
5      153
6      194
7      199
8      172
9      192
10     155
11     169
12     170
dtype: int64
```

```
In [278... sns.barplot(x=month.index,y=month)
plt.plot()
```

```
Out[278]: []
```

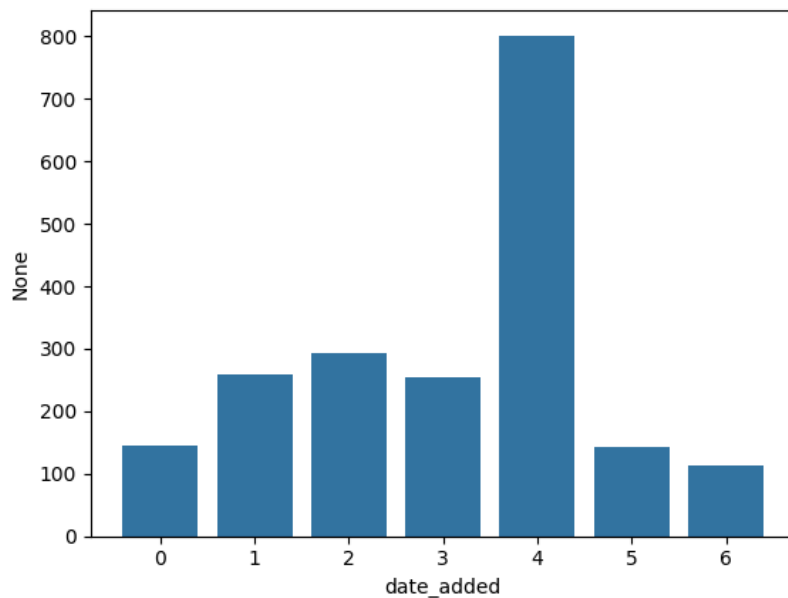


```
In [279... #as we are on thusday many movies is Launching
day=top_tv_shows.groupby(top_tv_shows["date_added"].dt.day_of_week).size()
day
```

```
Out[279]: date_added
0      144
1      258
2      293
3      255
4      802
5      143
6      114
dtype: int64
```

```
In [280... sns.barplot(x=day.index,y=day)
plt.plot()
```

```
Out[280]: []
```



Recommendation: Best time to release a Tv\_show on netflix on one of the month(june,july,aug,sep)

## The type of genre in which most Movies are released.

```
In [281... IDF["listed_in"].value_counts()
```

```
Out[281]: international movies    11229
dramas                        9591
comedies                      7033
international tv shows        6518
tv dramas                     3898
action & adventure            3353
tv comedies                   2955
children & family movies      2731
romantic movies               2727
independent movies            2376
crime tv shows                2375
thrillers                     2361
documentaries                 2337
kids' tv                      2278
romantic tv shows             1834
docuseries                    1782
music & musicals               1659
reality tv                    1552
horror movies                 1376
anime series                  1101
tv action & adventure          1004
spanish-language tv shows     977
stand-up comedy               957
british tv shows              871
sports movies                 758
korean tv shows               730
sci-fi & fantasy               684
tv mysteries                   601
anime features                521
tv sci-fi & fantasy            492
teen tv shows                 429
lgbtq movies                  399
tv horror                     390
classic movies                363
science & nature tv           342
tv thrillers                   338
faith & spirituality           194
stand-up comedy & talk shows   178
movies                        169
cult movies                   121
tv shows                      81
classic & cult tv              72
Name: listed_in, dtype: int64
```

International Movies Genre are More Popular or Produce.

## Analysis of Top Actors and Directors for different types of Movies and TV.

```
In [282... #Top 10 Directors appered in most movies
top_Mdirectors=IDF[["show_id", "type", "director"]].drop_duplicates(keep="first").reset_index(drop=True)
top_Mdirectors[top_Mdirectors["type"]=="Movie"]["director"].value_counts()[:10]
```

```
Out[282]: Martin Scorsese      200
          Rajiv Chilaka    19
          Raúl Campos, Jan Suter  18
          Suhas Kadav      16
          Marcus Raboy     15
          Jay Karas        14
          Cathy Garcia-Molina 13
          Youssef Chahine   12
          Jay Chapman       12
          Steven Spielberg  11
          Name: director, dtype: int64
```

These are the top 10 directors which appered in most movies.

```
In [283... #Top 10 Directors appered in most TV-shows
top_TVdirectors=IDF[["show_id","type","director"]].drop_duplicates(keep="first").reset_index(drop=True)
top_TVdirectors[top_TVdirectors["type"]=="TV Show"]["director"].value_counts()[ :10]
```

```
Out[283]: Martin Scorsese      2446
          Alastair Fothergill    3
          Hsu Fu-chun           2
          Iginio Straffi        2
          Shin Won-ho           2
          Rob Seidenglanz       2
          Ken Burns             2
          Stan Lathan           2
          Quek Shio-chuan       1
          Peter Ho              1
          Name: director, dtype: int64
```

These are the top 10 directors which appered in most TV Shows.

```
In [284... #Top 10 Cast appered in most Movies
top_MCast=IDF[["show_id","type","cast"]].drop_duplicates(keep="first").reset_index(drop=True)
top_MCast[top_MCast["type"]=="Movie"]["cast"].value_counts()[ :10]
```

```
Out[284]: anupam kher      517
          shah rukh khan  35
          naseeruddin shah 32
          om puri         30
          akshay kumar     30
          amitabh bachchan 28
          julie tejwani     28
          paresh rawal     28
          boman irani      27
          rupa bhimani      27
          Name: cast, dtype: int64
```

These are the top 10 Cast which appered in most Movies.

```
In [285... #Top 10 Cast appered in most TV-shows
top_TVCast=IDF[["show_id","type","cast"]].drop_duplicates(keep="first").reset_index(drop=True)
top_TVCast[top_TVCast["type"]=="TV Show"]["cast"].value_counts()[ :10]
```

```
Out[285]: anupam kher      351
          takahiro sakurai  25
          yuki kaji        19
          ai kayano        17
          daisuke ono       17
          junichi suwabe    17
          yuichi nakamura   16
          yoshimasa hosoya  15
          jun fukuyama      15
          david attenborough 14
          Name: cast, dtype: int64
```

These are the top 10 Cast which appered in most TV Shows.

```
In [286... IDF.head()
```

```
Out[286]:
```

	show_id	type	title	director	date_added	release_year	rating	duration	cast	country	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	September 25, 2021	2020	PG-13	90 min	anupam kher	united states	documentaries
1	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	ama qamata	south africa	international tv shows
2	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	khosi ngema	south africa	international tv shows
3	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	gail mabalane	south africa	international tv shows
4	s2	TV Show	Blood & Water	Martin Scorsese	September 24, 2021	2021	TV-MA	2 Seasons	thabang molaba	south africa	international tv shows

```
In [291... day_release=IDF[["show_id","type","release_year","date_added"]]
day_release["date_added"] = pd.to_datetime(day_release["date_added"])
```

```

day_release["date_added"] = day_release["date_added"].dt.year
day_release["diff"] = day_release["date_added"] - day_release["release_year"]
day_release.head()

```

```

<ipython-input-291-5def53534528>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```

day_release["date_added"] = pd.to_datetime(day_release["date_added"])

```

```

<ipython-input-291-5def53534528>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```

day_release["date_added"] = day_release["date_added"].dt.year

```

```

<ipython-input-291-5def53534528>:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```

day_release["diff"] = day_release["date_added"] - day_release["release_year"]

```

```

Out[291]:

```

	show_id	type	release_year	date_added	diff
0	s1	Movie	2020	2021	1
1	s2	TV Show	2021	2021	0
2	s2	TV Show	2021	2021	0
3	s2	TV Show	2021	2021	0
4	s2	TV Show	2021	2021	0

```

In [292... print("Mean=", day_release[day_release["type"]=="Movie"]["diff"].mean())
print("Mode=", day_release[day_release["type"]=="Movie"]["diff"].mode()[0])

```

```

Mean= 6.784357923497268
Mode= 0

```

In general the movies get added on Netflix within same year of release year.

```

In [294... print("Mean=", day_release[day_release["type"]=="TV Show"]["diff"].mean())
print("Mode=", day_release[day_release["type"]=="TV Show"]["diff"].mode()[0])

```

```

Mean= 2.523838287957889
Mode= 0

```

In general the TV-Shows get added on Netflix within same year of release year.

```

In [ ]:

```