NETFLIX-DATA EXPLORATION AND VISUALISATION

Netflix is widely recognized as a leading media and video streaming platform, renowned for its extensive library of over 10,000 movies and TV shows. As of mid-2021, Netflix boasts a staggering global subscriber base of over 222 million. The dataset in question comprises a tabular format that encompasses comprehensive listings of all the movies and TV shows accessible on the platform. It encompasses a wealth of details including cast members, directors, ratings, release years, durations, and more.

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

#Use the read_csv function to read the dataset from the specified file
path.
#Let's import and view the first five rows using the head() function.

#Reading dataset
df=pd.read_csv("Netflix case study dataset.csv")
df.head()
```

```
  show_id     type                   title         director  \
0      s1    Movie   Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show          Blood & Water              NaN
2      s3  TV Show              Ganglands  Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans              NaN
4      s5  TV Show           Kota Factory              NaN

                                                cast        country  \
0                                                NaN  United States
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...   South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...            NaN
3                                                NaN            NaN
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...          India

          date_added  release_year rating   duration  \
0  September 25, 2021          2020  PG-13     90 min
1  September 24, 2021          2021  TV-MA  2 Seasons
2  September 24, 2021          2021  TV-MA   1 Season
3  September 24, 2021          2021  TV-MA   1 Season
4  September 24, 2021          2021  TV-MA  2 Seasons

                                        listed_in  \
0                                   Documentaries
1    International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act...
3                            Docuseries, Reality TV
```

```
4  International TV Shows, Romantic TV Shows, TV ...

                                          description
0  As her father nears the end of his life, filmm...
1  After crossing paths at a party, a Cape Town t...
2  To protect his family from a powerful drug lor...
3  Feuds, flirtations and toilet talk go down amo...
4  In a city of coaching centers known to train I...
```

Observations on the shape of data, data types of all the attributes and exploring the data

```
#Checking the number of rows and columns
df.shape
```

```
(8807, 12)
```

```
#To list the names of columns present
df.columns
```

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

```
#To check the datatype of the columns
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
```

BASIC ANALYSIS OF DATASET

```
#Checking the null value
df1.isnull().sum()
```

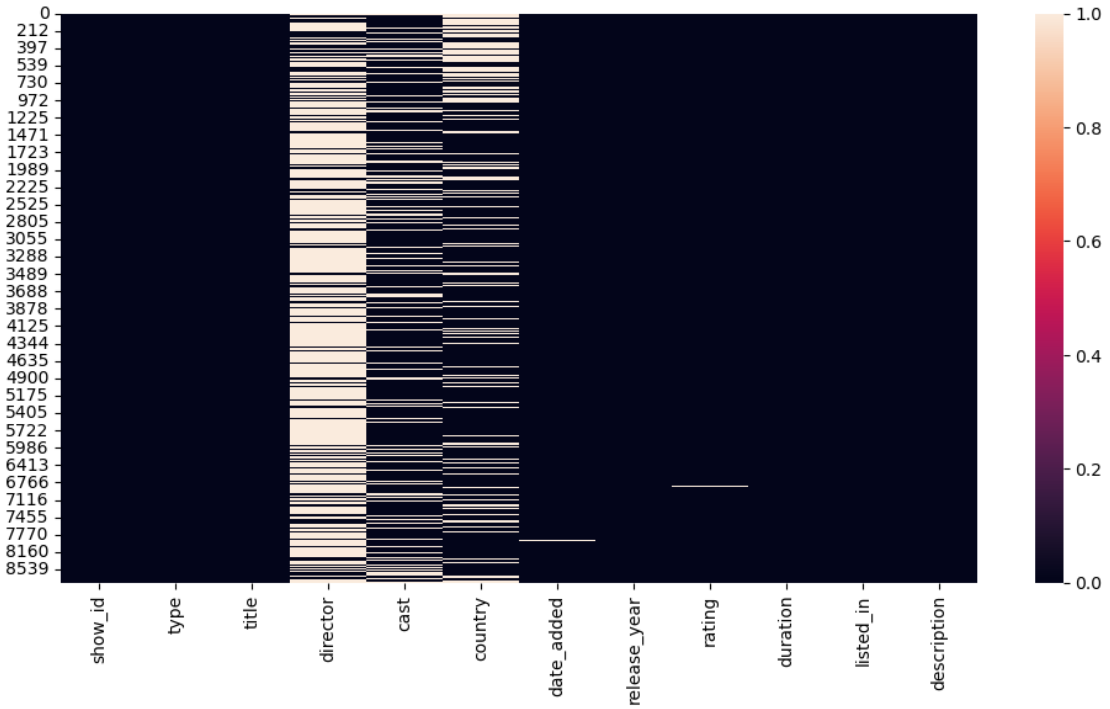
```
show_id            0
type               0
title              0
director        2634
cast             825
country          831
date_added        10
release_year       0
rating             4
duration           0
listed_in          0
description        0
dtype: int64
```

```
df1=df1[df1.isnull().any(axis=1)]
```

```
#Heatmap showing null values

plt.figure(figsize=[12,6])
sns.heatmap(df1.isnull())
```

```
<Axes: >
```



OBSERVATION : NULL VALUES ARE PRESENT IN director,cast,country,date_added,rating,duration COLUMNS

1. Drop NaN Values & handling null values

```
df1=df
df1=df.dropna(subset=["rating"])
df1=df.dropna(subset=["duration"])
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8804 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8804 non-null   object
 1   type          8804 non-null   object
 2   title         8804 non-null   object
 3   director      6170 non-null   object
 4   cast          7979 non-null   object
 5   country       7973 non-null   object
 6   date_added    8794 non-null   object
 7   release_year  8804 non-null   int64 
 8   rating        8800 non-null   object
 9   duration      8804 non-null   object
 10  listed_in     8804 non-null   object
 11  description   8804 non-null   object
dtypes: int64(1), object(11)
memory usage: 894.2+ KB
```

```
df_null=df1[["country","cast","director"]].isna()
df_null=df_null[(df_null["country"]==True) & (df_null["cast"]==True) &
(df_null["director"]==True)]
index_to_drop=df_null.index
index_to_drop
```

```
Int64Index([   3,   10,   14,   74,  123,  147,  181,  218,  224,
234,  242,
             320,  368,  369,  379,  404,  430,  465,  466,  490,
497,  549,
             641,  653,  710,  738,  761,  842,  851,  852,  854,
925,  936,
             957, 1005, 1038, 1056, 1094, 1095, 1116, 1224, 1225,
1723, 1828,
            1889, 1890, 1952, 2101, 2152, 2222, 2236, 2335, 2394,
2425, 2432,
            2495, 2568, 2569, 2699, 2893, 2982, 3393, 3394, 3395,
3396, 3397,
            3469, 3539, 3592, 3789, 3837, 3953, 3987, 4015, 4056,
4199, 4250,
            4372, 4627, 4637, 4654, 5291, 5799, 5870, 6389, 6618,
7017, 7222,
            7438, 7483, 7624, 7812, 8109, 8199, 8609, 8803],
           dtype='int64')
```

```
df1=df1.drop(index=index_to_drop)
df1
```

```
     show_id     type                  title         director  \
0         s1    Movie  Dick Johnson Is Dead  Kirsten Johnson
1         s2  TV Show         Blood & Water              NaN
2         s3  TV Show             Ganglands  Julien Leclercq
4         s5  TV Show          Kota Factory              NaN
5         s6  TV Show         Midnight Mass    Mike Flanagan
...      ...      ...                   ...              ...
8801   s8802    Movie               Zinzana  Majid Al Ansari
8802   s8803    Movie                Zodiac    David Fincher
8804   s8805    Movie            Zombieland  Ruben Fleischer
8805   s8806    Movie                  Zoom     Peter Hewitt
8806   s8807    Movie                Zubaan      Mozez Singh

                                                   cast  \
0                                                   NaN
1     Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...
2     Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...
4     Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...
5     Kate Siegel, Zach Gilford, Hamish Linklater, H...
...                                                 ...
8801  Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...
8802  Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...
8804  Jesse Eisenberg, Woody Harrelson, Emma Stone, ...
8805  Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...
8806  Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...

                            country          date_added  release_year
rating  \
0                     United States  September 25, 2021          2020
PG-13
1                      South Africa  September 24, 2021          2021
TV-MA
2                               NaN  September 24, 2021          2021
TV-MA
4                             India  September 24, 2021          2021
TV-MA
5                               NaN  September 24, 2021          2021
TV-MA
...                             ...                 ...           ...
...
8801  United Arab Emirates, Jordan       March 9, 2016          2015
TV-MA
8802                 United States   November 20, 2019          2007
R
8804                 United States    November 1, 2019          2009
R
8805                 United States    January 11, 2020          2006
```

```
PG
8806                         India       March 2, 2019          2015
TV-14

       duration                                                listed_in  \
0        90 min                                          Documentaries
1     2 Seasons    International TV Shows, TV Dramas, TV Mysteries
2      1 Season  Crime TV Shows, International TV Shows, TV Act...
4     2 Seasons  International TV Shows, Romantic TV Shows, TV ...
5      1 Season                 TV Dramas, TV Horror, TV Mysteries
...         ...                                                 ...
8801     96 min           Dramas, International Movies, Thrillers
8802    158 min                     Cult Movies, Dramas, Thrillers
8804     88 min                           Comedies, Horror Movies
8805     88 min                 Children & Family Movies, Comedies
8806    111 min     Dramas, International Movies, Music & Musicals

                                              description
0     As her father nears the end of his life, filmm...
1     After crossing paths at a party, a Cape Town t...
2     To protect his family from a powerful drug lor...
4     In a city of coaching centers known to train I...
5     The arrival of a charismatic young priest brin...
...                                                 ...
8801  Recovering alcoholic Talal wakes up inside a s...
8802  A political cartoonist, a crime reporter and a...
8804  Looking to survive in a world taken over by zo...
8805  Dragged from civilian life, a former superhero...
8806  A scrappy but poor boy worms his way into a ty...

[8708 rows x 12 columns]
```

IMPUTATION OF DATE_ADDED

```
df1["date_added"]=df1["date_added"].fillna(df1["release_year"])
df1["date_added"].isna().sum()
```

```
0
```

IMPUTATION OF DIRECTORS

```
director_fill=df1.groupby(["type","listed_in","country"])
["director"].apply(lambda x:x.mode()) #filling the data with value
that appears most
director_fill=director_fill.reset_index()
director_fill
```

```
         type           listed_in                   country
level_3  \
```

```
0       Movie  Action & Adventure  Australia, United States        0
1       Movie  Action & Adventure   Bulgaria, United States        0
2       Movie  Action & Adventure                    Canada        0
3       Movie  Action & Adventure                    Canada        1
4       Movie  Action & Adventure                    Canada        2
...       ...                 ...                       ...      ...
3196  TV Show            TV Shows                     Japan        0
3197  TV Show            TV Shows              Saudi Arabia        0
3198  TV Show            TV Shows             United States        0
3199  TV Show            TV Shows             United States        1
3200  TV Show            TV Shows             United States        2


                                              director
0                                          Simon Wincer
1                                        Todor Chapkanov
2                                            Adam Alleca
3                                            Allan Ungar
4                                          Brett Donowho
...                                                  ...
3196                     Kobun Shizuno, Hiroyuki Seshita
3197  Ali Kalthami, Meshal Aljaser, Faris Godus, Moh...
3198                       Joe Berlinger, Bruce Sinofsky
3199  Oscar Micheaux, Spencer Williams, Richard E. N...
3200                       Steven Bognar, Julia Reichert

[3201 rows x 5 columns]
```

```
df1=df1.merge(director_fill,on=["type","listed_in","country"],how="lef
t")
df1["director"]=np.where(df1["director_x"].isna(),df1["director_y"],df
1["director_x"])
df1=df1[["show_id","type","title","director","cast","country","date_ad
ded","release_year","rating","duration","listed_in","description"]]
df1=df1.drop_duplicates(subset="show_id")
df1.head()
```

```
  show_id     type                  title         director  \
0      s1    Movie  Dick Johnson Is Dead  Kirsten Johnson
```

```
1      s2  TV Show         Blood & Water              NaN
2      s3  TV Show             Ganglands  Julien Leclercq
3      s5  TV Show          Kota Factory              NaN
4      s6  TV Show         Midnight Mass    Mike Flanagan

                                                  cast        country  \
0                                                  NaN  United States
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...   South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...            NaN
3  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...          India
4  Kate Siegel, Zach Gilford, Hamish Linklater, H...            NaN

           date_added  release_year rating   duration  \
0  September 25, 2021          2020  PG-13     90 min
1  September 24, 2021          2021  TV-MA  2 Seasons
2  September 24, 2021          2021  TV-MA   1 Season
3  September 24, 2021          2021  TV-MA  2 Seasons
4  September 24, 2021          2021  TV-MA   1 Season

                                        listed_in  \
0                                   Documentaries
1    International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act...
3  International TV Shows, Romantic TV Shows, TV ...
4                TV Dramas, TV Horror, TV Mysteries

                                         description
0  As her father nears the end of his life, filmm...
1  After crossing paths at a party, a Cape Town t...
2  To protect his family from a powerful drug lor...
3  In a city of coaching centers known to train I...
4  The arrival of a charismatic young priest brin...
```

```
df2=df1.merge(director_fill[["type","listed_in","director"]],on=["type
","listed_in"],how="left")
df2["director"]=np.where(df2["director_x"].isna(),df2["director_y"],df
2["director_x"])
df2=df2[["show_id","type","title","director","cast","country","date_ad
ded","release_year","rating","duration","listed_in","description"]]
df2=df2.drop_duplicates(subset="show_id")
df2
```

```
       show_id     type                  title           director  \
0           s1    Movie  Dick Johnson Is Dead     Kirsten Johnson
48          s2  TV Show         Blood & Water  Michel Tikhomiroff
52          s3  TV Show             Ganglands     Julien Leclercq
54          s5  TV Show          Kota Factory           BB Sasore
60          s6  TV Show         Midnight Mass       Mike Flanagan
...        ...      ...                   ...                 ...
357635   s8802    Movie               Zinzana     Majid Al Ansari
```

```
357749   s8803    Movie              Zodiac       David Fincher
357750   s8805    Movie           Zombieland     Ruben Fleischer
357754   s8806    Movie                Zoom        Peter Hewitt
357834   s8807    Movie              Zubaan         Mozez Singh

                                                        cast  \
0                                                        NaN
48      Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...
52      Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...
54      Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...
60      Kate Siegel, Zach Gilford, Hamish Linklater, H...
...                                                    ...
357635  Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...
357749  Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...
357750  Jesse Eisenberg, Woody Harrelson, Emma Stone, ...
357754  Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...
357834  Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...

                             country          date_added  release_year
rating  \
0                      United States  September 25, 2021          2020
PG-13
48                      South Africa  September 24, 2021          2021
TV-MA
52                               NaN  September 24, 2021          2021
TV-MA
54                             India  September 24, 2021          2021
TV-MA
60                               NaN  September 24, 2021          2021
TV-MA
...                              ...                 ...           ...
...
357635  United Arab Emirates, Jordan       March 9, 2016          2015
TV-MA
357749                 United States   November 20, 2019          2007
R
357750                 United States    November 1, 2019          2009
R
357754                 United States    January 11, 2020          2006
PG
357834                         India       March 2, 2019          2015
TV-14

         duration
listed_in  \
0          90 min                                    Documentaries

48      2 Seasons    International TV Shows, TV Dramas, TV Mysteries

52       1 Season  Crime TV Shows, International TV Shows, TV Act...
```

```
54      2 Seasons  International TV Shows, Romantic TV Shows, TV ...

60       1 Season                TV Dramas, TV Horror, TV Mysteries

...           ...                                              ...

357635     96 min            Dramas, International Movies, Thrillers

357749    158 min                    Cult Movies, Dramas, Thrillers

357750     88 min                           Comedies, Horror Movies

357754     88 min                Children & Family Movies, Comedies

357834    111 min     Dramas, International Movies, Music & Musicals


                                                 description
0       As her father nears the end of his life, filmm...
48      After crossing paths at a party, a Cape Town t...
52      To protect his family from a powerful drug lor...
54      In a city of coaching centers known to train I...
60      The arrival of a charismatic young priest brin...
...                                                   ...
357635  Recovering alcoholic Talal wakes up inside a s...
357749  A political cartoonist, a crime reporter and a...
357750  Looking to survive in a world taken over by zo...
357754  Dragged from civilian life, a former superhero...
357834  A scrappy but poor boy worms his way into a ty...

[8708 rows x 12 columns]
```

```
df3=df2.merge(director_fill[["type","director"]],on=["type"],how="left
")
df3["director"]=np.where(df3["director_x"].isna(),df3["director_y"],df
3["director_x"])
df3=df3[["show_id","type","title","director","cast","country","date_ad
ded","release_year","rating","duration","listed_in","description"]]
df3=df3.drop_duplicates(subset="show_id")

df3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8708 entries, 0 to 18971439
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       8708 non-null   object
```

```
 1   type          8708 non-null   object
 2   title         8708 non-null   object
 3   director      8708 non-null   object
 4   cast          7979 non-null   object
 5   country       7973 non-null   object
 6   date_added    8708 non-null   object
 7   release_year  8708 non-null   int64
 8   rating        8704 non-null   object
 9   duration      8708 non-null   object
 10  listed_in     8708 non-null   object
 11  description   8708 non-null   object
dtypes: int64(1), object(11)
memory usage: 884.4+ KB
```

IMPUTATION OF COUNTRY

```
country_fill=df1.groupby(["cast","director"])["country"].apply(lambda
x:x.mode())
country_fill=country_fill.reset_index()
country_fill
```

```
                                                   cast  \
0     50 Cent, Ryan Phillippe, Bruce Willis, Rory Ma...
1     A.J. LoCascio, Sendhil Ramamurthy, Fred Tatasc...
2                                            A.R. Rahman
3     Aadhi, Tapsee Pannu, Ritika Singh, Vennela Kis...
4     Aadil Khan, Sadia Khateeb, Zain Khan Durrani, ...
...                                                 ...
6035  Úrsula Corberó, Álvaro Cervantes, Najwa Nimri,...
6036  İbrahim Büyükak, Zeynep Koçak, Gupse Özay, Cen...
6037  Şahin Irmak, İrem Sak, Gonca Vuslateri, Emre K...
6038  Şükrü Özyıldız, Aslı Enver, Şenay Gürler, Başa...
6039  Ṣọpẹ́ Dìrísù, Wunmi Mosaku, Matt Smith, Malaik...

                     director  level_2         country
0                Mike Gunther         0   United States
1             Michael D. Black        0   United States
2     Rudradeep Bhattacharjee         0           India
3                   Hari Nath         0           India
4          Vidhu Vinod Chopra         0           India
...                         ...      ...             ...
6035               Julio Medem         0           Spain
6036              Bedran Güzel         0          Turkey
6037               Hakan Algül         0          Turkey
6038       Neslihan Yesilyurt          0          Turkey
6039                Remi Weekes        0  United Kingdom

[6040 rows x 4 columns]
```

```
df1=df1.merge(country_fill,on=["cast","director"],how="left")
df1["country"]=np.where(df1["country_x"].isna(),df1["country_y"],df1["
country_x"])
df1=df1[["show_id","type","title","director","cast","country","date_ad
ded","release_year","rating","duration","listed_in","description"]]
df1=df1.drop_duplicates(subset="show_id")

df2=df1.merge(country_fill[["country","cast"]],on=["cast"],how="left")
df2["country"]=np.where(df2["country_x"].isna(),df2["country_y"],df2["
country_x"])
df2=df2[["show_id","type","title","director","cast","country","date_ad
ded","release_year","rating","duration","listed_in","description"]]
df2=df2.drop_duplicates(subset="show_id")

df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8708 entries, 0 to 9164
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8708 non-null   object
 1   type          8708 non-null   object
 2   title         8708 non-null   object
 3   director      7068 non-null   object
 4   cast          7979 non-null   object
 5   country       8034 non-null   object
 6   date_added    8708 non-null   object
 7   release_year  8708 non-null   int64 
 8   rating        8704 non-null   object
 9   duration      8708 non-null   object
 10  listed_in     8708 non-null   object
 11  description   8708 non-null   object
dtypes: int64(1), object(11)
memory usage: 884.4+ KB
```

```
Not_available=df["country"].mode()[0]
df2["country"]=df2["country"].fillna(Not_available)
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8708 entries, 0 to 9164
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8708 non-null   object
 1   type          8708 non-null   object
 2   title         8708 non-null   object
 3   director      7068 non-null   object
 4   cast          7979 non-null   object
 5   country       8708 non-null   object
```

```
 6   date_added    8708 non-null   object
 7   release_year  8708 non-null   int64
 8   rating        8704 non-null   object
 9   duration      8708 non-null   object
 10  listed_in     8708 non-null   object
 11  description   8708 non-null   object
dtypes: int64(1), object(11)
memory usage: 884.4+ KB
```

CONVERTING OBJECT TO DATETIME

```
df2["date_added"]=pd.to_datetime(df2["date_added"])
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8708 entries, 0 to 9164
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       8708 non-null   object
 1   type          8708 non-null   object
 2   title         8708 non-null   object
 3   director      7068 non-null   object
 4   cast          7979 non-null   object
 5   country       8708 non-null   object
 6   date_added    8708 non-null   datetime64[ns]
 7   release_year  8708 non-null   int64
 8   rating        8704 non-null   object
 9   duration      8708 non-null   object
 10  listed_in     8708 non-null   object
 11  description   8708 non-null   object
dtypes: datetime64[ns](1), int64(1), object(10)
memory usage: 884.4+ KB
```

2. Un-nesting the columns and handling the missing values

```
#Unnesting cast,director,country,listed_in(genre)
```

- Unnesting directors

```
unnest_column=df1["director"].str.split(",",expand=True).add_prefix("director")
```

```
df2=df1.merge(unnest_column,left_index=True,right_index=True)
df2.drop("director",axis=1,inplace=True)  # To drop original director's column
```

```
df3=pd.melt(df2,id_vars=["show_id","type","title","cast","country","date_added","release_year","rating","duration","listed_in","description"]
                        ,value_name="director").drop("variable",axis=1)
```

```
df3.dropna(subset=["director"],inplace=True)   #Dropping None values
```

```
df1=df3
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 159527 entries, 0 to 159526
Data columns (total 12 columns):
 #   Column        Non-Null Count   Dtype
---  ------        --------------   -----
 0   show_id       159527 non-null  object
 1   type          159527 non-null  object
 2   title         159527 non-null  object
 3   cast          159527 non-null  object
 4   country       159527 non-null  object
 5   date_added    159527 non-null  object
 6   release_year  159527 non-null  int64
 7   rating        159476 non-null  object
 8   duration      159527 non-null  object
 9   listed_in     159527 non-null  object
 10  description   159527 non-null  object
 11  director      159527 non-null  object
dtypes: int64(1), object(11)
memory usage: 14.6+ MB
```

```
df1.head(3)
```

```
  show_id     type                          title         director  \
0      s1    Movie           Dick Johnson Is Dead  Kirsten Johnson
4      s8    Movie                         Sankofa     Haile Gerima
5      s9  TV Show  The Great British Baking Show  Andy Devonshire

                                                  cast  \
0                                                  NaN
4  Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
5  Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...

                                       listed_in
date_added  \
0                                  Documentaries  September 25,
2021
4  Dramas, Independent Movies, International Movies  September 24,
2021
5                    British TV Shows, Reality TV  September 24,
2021

   release_year rating   duration  \
0          2020  PG-13     90 min
4          1993  TV-MA    125 min
5          2021  TV-14  9 Seasons

                                        description         country
```

```
0  As her father nears the end of his life, filmm...   United States
4  On a photo shoot in Ghana, an American model s...   United States
5  A talented batch of amateur bakers face off in...  United Kingdom
```

- Unnesting listed_in(Genre)

```
unnest_column=df1["listed_in"].str.split(",",expand=True).add_prefix("
listed_in")
df2=df1.merge(unnest_column,left_index=True,right_index=True)
df2.drop("listed_in",axis=1,inplace=True)
df3=pd.melt(df2,id_vars=["show_id","type","title","director","cast","c
ountry","date_added","release_year","rating","duration","description"]
                        ,value_name="listed_in").drop("variable",axis=
1)

df3.dropna(subset=["listed_in"],inplace=True)
df1=df3
df1.head(3)
```

```
  show_id     type                          title         director  \
0      s8    Movie                         Sankofa     Haile Gerima
1      s9  TV Show  The Great British Baking Show  Andy Devonshire
2     s10    Movie                    The Starling   Theodore Melfi

                cast         country          date_added  release_year
rating  \
0      Kofi Ghanaba   United States  September 24, 2021          1993
TV-MA
1      Mel Giedroyc  United Kingdom  September 24, 2021          2021
TV-14
2  Melissa McCarthy   United States  September 24, 2021          2021
PG-13

    duration                                           description  \
0    125 min  On a photo shoot in Ghana, an American model s...
1  9 Seasons  A talented batch of amateur bakers face off in...
2    104 min  A woman adjusting to life after a loss contend...

          listed_in
0            Dramas
1  British TV Shows
2          Comedies
```

- Unnesting country

```
unnest_column=df1["country"].str.split(",",expand=True).add_prefix("co
untry")
df2=df1.merge(unnest_column,left_index=True,right_index=True)
df2.drop("country",axis=1,inplace=True)
df3=pd.melt(df2,id_vars=["show_id","type","title","director","cast","l
isted_in","date_added","release_year","rating","duration","description
"]
```

```
                        ,value_name="country").drop("variable",axis=1)


df3.dropna(subset=["country"],inplace=True)
df1=df3
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 159527 entries, 0 to 159526
Data columns (total 12 columns):
 #   Column        Non-Null Count   Dtype 
---  ------        --------------   ----- 
 0   show_id       159527 non-null  object
 1   type          159527 non-null  object
 2   title         159527 non-null  object
 3   director      159527 non-null  object
 4   cast          159527 non-null  object
 5   listed_in     159527 non-null  object
 6   date_added    159527 non-null  object
 7   release_year  159527 non-null  int64 
 8   rating        159476 non-null  object
 9   duration      159527 non-null  object
 10  description   159527 non-null  object
 11  country       159527 non-null  object
dtypes: int64(1), object(11)
memory usage: 14.6+ MB
```

- Unnesting cast

```
unnest_column=df1["cast"].str.split(",",expand=True).add_prefix("cast"
)
df2=df1.merge(unnest_column,left_index=True,right_index=True)
df2.drop("cast",axis=1,inplace=True)
df3=pd.melt(df2,id_vars=["show_id","type","title","director","listed_i
n","country","date_added","release_year","rating","duration","descript
ion"]
                        ,value_name="cast").drop("variable",axis=1)

df3.dropna(subset=["cast"],inplace=True)
df1=df3
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 159527 entries, 0 to 159526
Data columns (total 12 columns):
 #   Column        Non-Null Count   Dtype 
---  ------        --------------   ----- 
 0   show_id       159527 non-null  object
 1   type          159527 non-null  object
 2   title         159527 non-null  object
 3   director      159527 non-null  object
```

```
 4   listed_in     159527 non-null  object
 5   country       159527 non-null  object
 6   date_added    159527 non-null  object
 7   release_year  159527 non-null  int64
 8   rating        159476 non-null  object
 9   duration      159527 non-null  object
 10  description   159527 non-null  object
 11  cast          159527 non-null  object
dtypes: int64(1), object(11)
memory usage: 14.6+ MB
```

```
df1["country"].value_counts().head(10)
```

```
United States     22365
India              8277
 United States     3600
United Kingdom     3511
Japan              2215
Canada             1620
 France            1572
South Korea        1500
 Canada            1409
Spain              1390
Name: country, dtype: int64
```

Unnesting cast,director,country,listed_in is done Handling none values and imputaion is also done succesfully

Analysis

Business problem : Analyze the data and generate insights that could help Netflix ijn deciding which type of shows/movies to produce and how they can grow the business in different countries

Q1: *Find the counts of each categorical variable both using graphical and non-graphical analysis.*

```
#Non graphical
```

```
df1["type"].value_counts()
```

```
Movie      138800
TV Show     20727
Name: type, dtype: int64
```

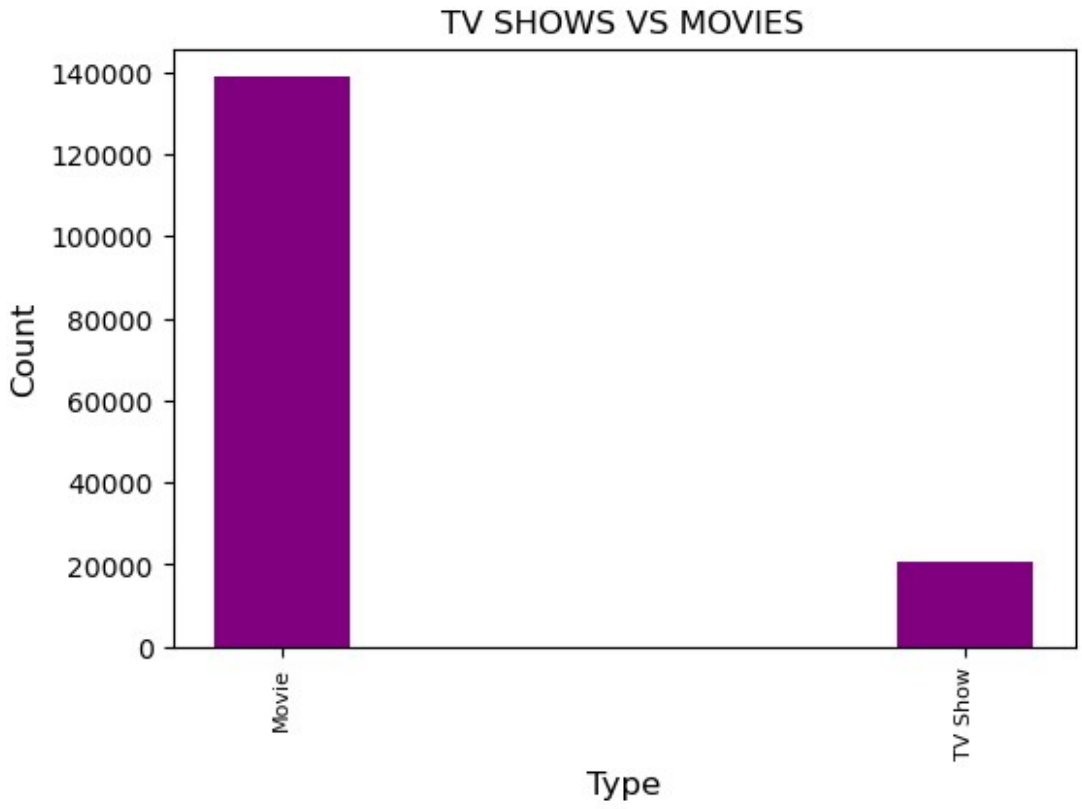
```
type_counts=df1["type"].value_counts()
type_counts
```

```
Movie      138800
TV Show     20727
Name: type, dtype: int64
```

```
x=type_counts.index
y=type_counts

#Graphical

plt.figure(figsize=[6,4])
plt.bar(x,y,width=0.2,color="purple")
plt.xticks(rotation = 90, fontsize = 8)
plt.xlabel("Type",fontsize=12)
plt.ylabel("Count",fontsize=12)
plt.title("TV SHOWS VS MOVIES")
plt.show()
```




```
df1["type"].value_counts()
```

```
Movie      138800
TV Show     20727
Name: type, dtype: int64
```

```
df1["director"].value_counts()
```

```
Hayato Date        1846
Shin Won-ho        1320
Felipe Cano         744
Kenny Ortega        646
Hsu Fu-chun         612
                   ...
Milton Horowitz       1
Matthew Cooke         1
Max Amini             1
Jason Orley           1
Ben Simms             1
Name: director, Length: 4420, dtype: int64
```

```
df1["country"].value_counts()
```

```
United States     42036
India             22090
United Kingdom     7925
 United States     7720
Japan              5126
                   ...
 Panama               2
 Sri Lanka            2
 Afghanistan          2
Venezuela             2
 Nicaragua            1
Name: country, Length: 175, dtype: int64
```

```
df1["listed_in"].value_counts()
```

```
 International Movies    25411
Dramas                   19022
Comedies                 13344
Action & Adventure       11834
 Dramas                   9876
                         ...
Classic & Cult TV           18
 Reality TV                 15
Romantic TV Shows           11
 Science & Nature TV        11
LGBTQ Movies                 5
Name: listed_in, Length: 70, dtype: int64
```

- Comparison of tv shows vs. movies¶

Q2: *How many movies and shows are in the dataset? Show with the bar graph.*

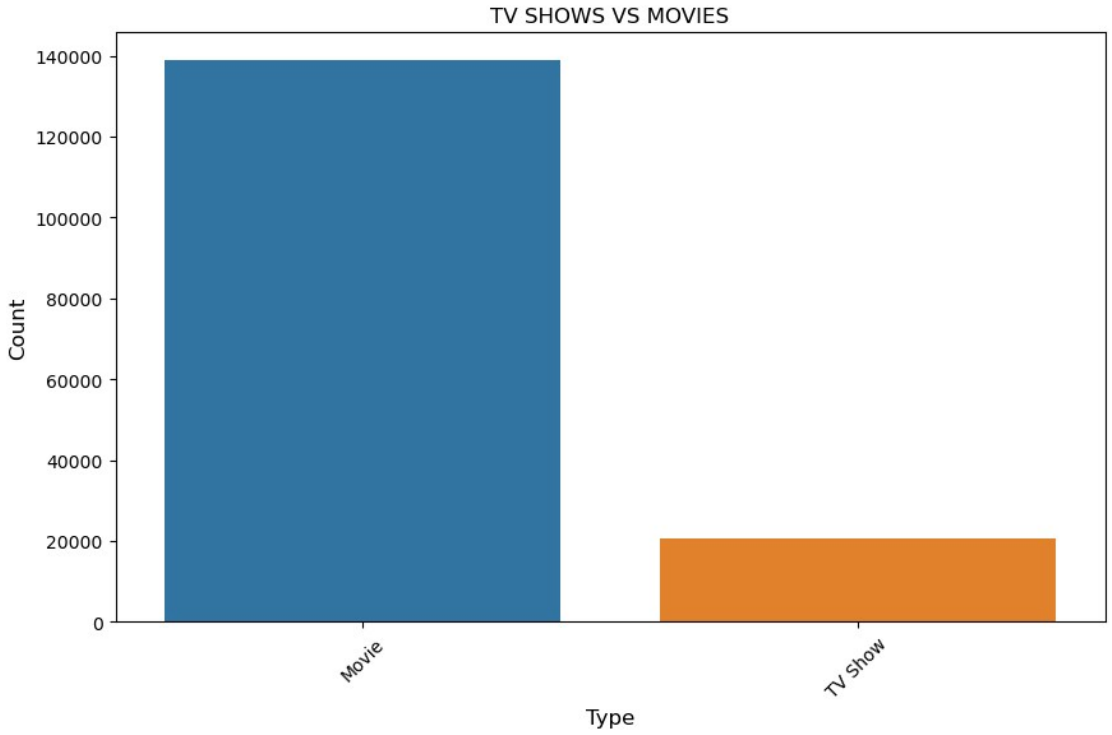
```
df1["type"].value_counts()
```

```
Movie      138800
TV Show     20727
Name: type, dtype: int64
```

```
#USING SEABORN
plt.figure(figsize=[10,6])
sns.countplot(data=df1,
              x="type"),


plt.xticks(rotation=45,fontsize=10)
plt.xlabel("Type",fontsize=12)
plt.ylabel("Count",fontsize=12)
plt.title("TV SHOWS VS MOVIES")

Text(0.5, 1.0, 'TV SHOWS VS MOVIES')
```



Business Insights

- We can clearly see that Movies are the released in huge numbers than TV Shows which indicates that audience enjoy movies than TV Shows may be due to the less duration in movies

Recommendations

- We should add more number of TV Shows to the platform to increase its viewership as well
- More and more movies can be released as it has high audience

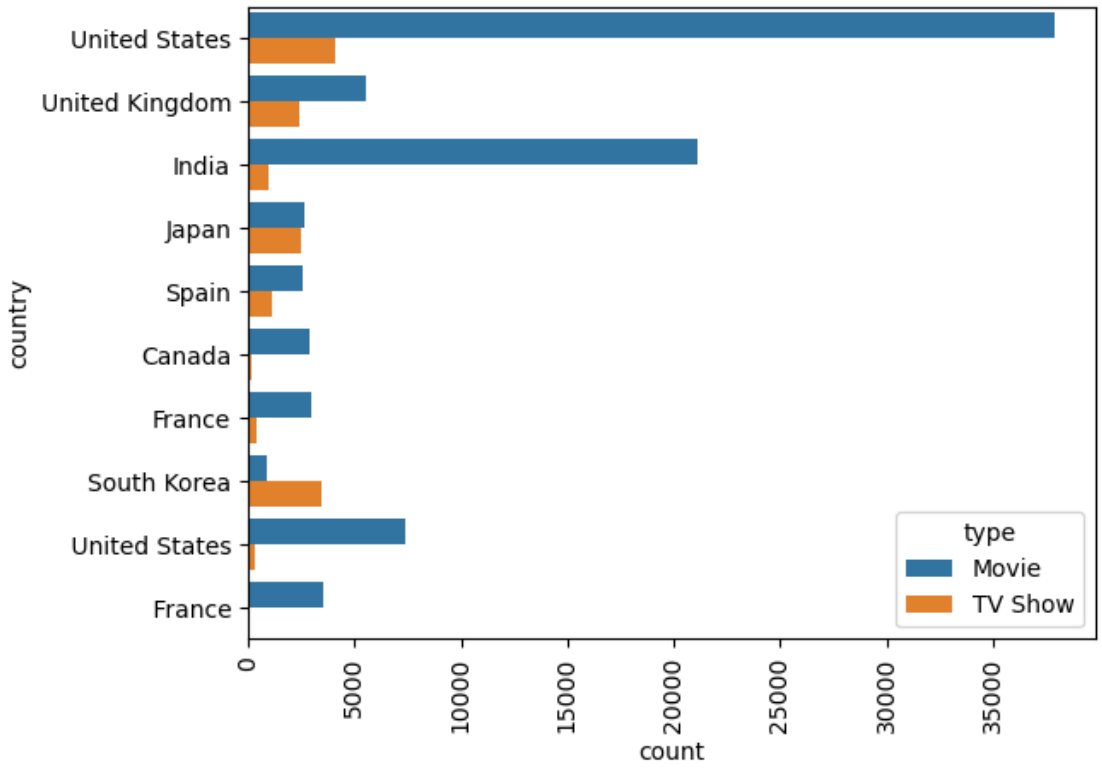
Q3: *What type of content is available in various countries?*.

```
countries=df1["country"].value_counts().head(10)
countries=countries.reset_index()
countries["index"]
```

```
0     United States
1             India
2    United Kingdom
3     United States
4             Japan
5       South Korea
6             Spain
7            France
8            France
9            Canada
Name: index, dtype: object
```

```
data=df1[df1["country"].isin(countries["index"])]

sns.countplot(data=data,y="country",hue="type")
plt.xticks(rotation=90)
plt.show()
```



Business Insights

- After analysing the top 10 countries,the Dodged bar chart shows that US and India tops the market for Movies.

- US tops the market for TV Shows . -Compared to other European counties UK holds the top markets for both movies and TV Shows. -There are hardly any TV Shows coming from Canada and France -Japan and South Korea has more viewership in TV Shows than in Movies.

Recommendations

- In India and US we can increase the number of movies produced per year as it has large audience
- We can produce more TV Shows in Japan and South Korea as it has more viewers

Q4: *In which year highest number of TV Shows and Movies were released? Show with bar graph.*

```
year_counts=df1["release_year"].value_counts()
year_counts
```

```
2018    18652
2017    16737
2016    15411
2019    14901
2020    12236
        ...
1959       12
1945        8
1947        8
1942        4
1946        2
Name: release_year, Length: 72, dtype: int64
```

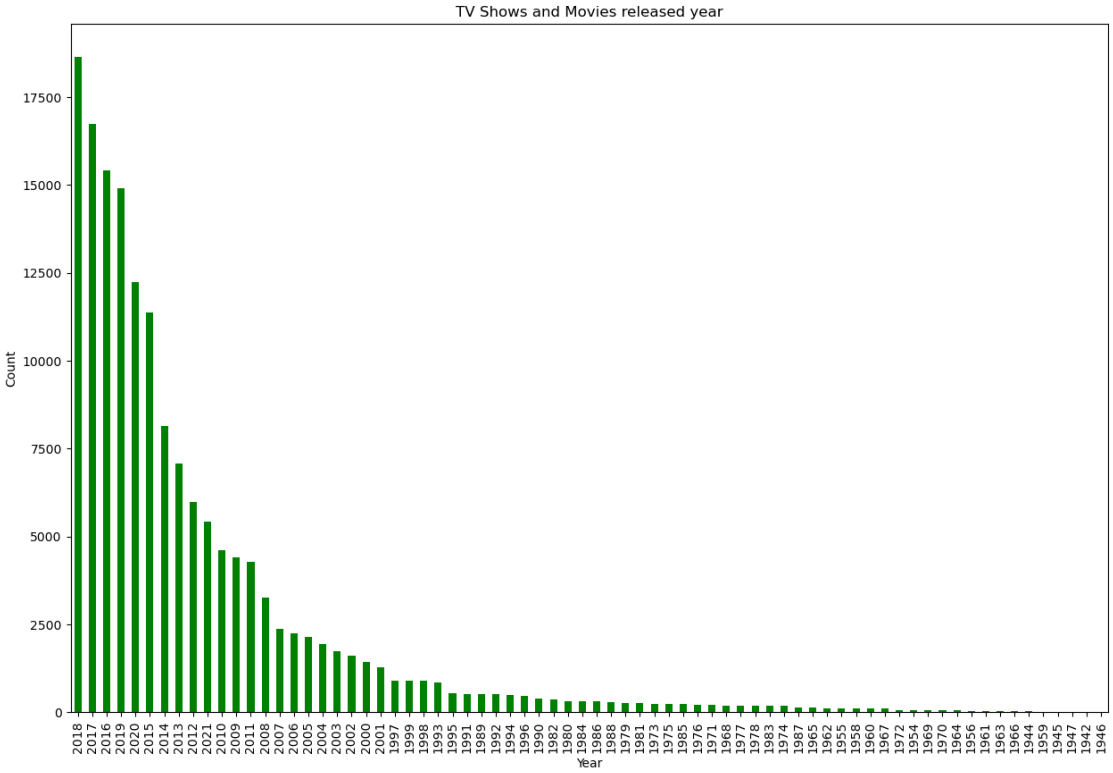
```
x=year_counts.index
y=year_counts
```

```
y_list=list(df1["release_year"].value_counts().index)
```

```
df1_sorted=y_list.sort()
```

```
#Show with bar graph

plt.figure(figsize=[15,10])
year_counts.plot(kind="bar",width=0.5,color="green")
plt.xlabel("Year",fontsize=10)
plt.ylabel("Count",fontsize=10)
plt.title("TV Shows and Movies released year")
plt.show()
```



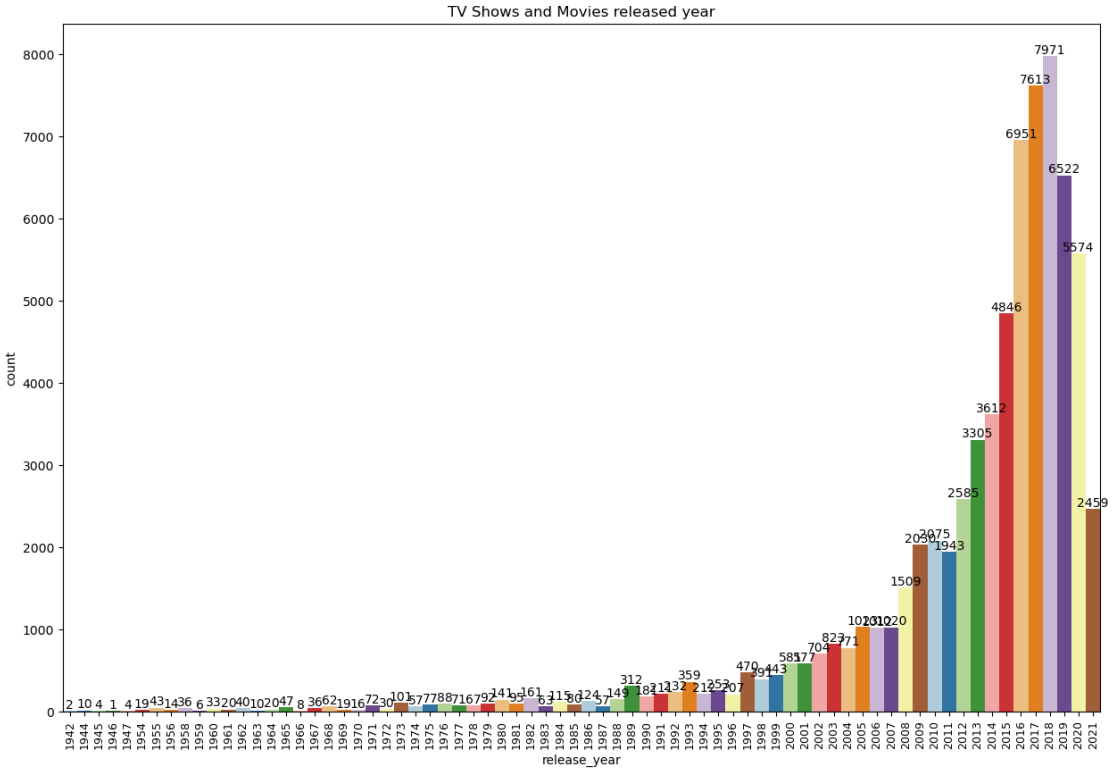

```
plt.figure (figsize=(15,10))

ax=sns.countplot(data=df1,
                 x="release_year",
                order=y_list,
                width=1,
                palette="Paired")

#ax containers

for i in ax.containers:
    ax.bar_label(i)

plt.xticks(rotation=90,fontsize=9)
plt.title("TV Shows and Movies released year")
plt.show()
```



OBSERVATION

- 2018 HAS HIGHEST NUMBER OF TV SHOWS AND MOVIES RELEASED (COUNT=7971) AND SHOWED THE HIGHEST PEAK FOLLOWING WHICH THERE HAS BEEN A DECREASE IN THE MOVIES AND SHOWS RELEASED

Q5: *Show the top 5 movies released based on their year of release*

```
movie_data=df1.loc[(df1["type"])=="Movie"]
sorted_movie_data=movie_data.sort_values(["release_year"],ascending=Fa
lse)
sorted_movie_data.head()
```

```
       show_id   type           title                    director  \
146362    s769  Movie  Dancing Queens           Helena Bergström
31600     s271  Movie         Beckett  Ferdinando Cito Filomarino
31621     s308  Movie       Aftermath               Peter Winther
97203     s878  Movie    Cinema Bandi         Praveen Kandregula
97202     s874  Movie           Ahaan            Nikhil Pherwani

                     listed_in        country       date_added
release_year  \
146362   International Movies         Sweden     June 3, 2021
2021
31600      Action & Adventure          Italy  August 13, 2021
2021
31621           Horror Movies  United States   August 4, 2021
```

```
2021
97203      Independent Movies          India     May 14, 2021
2021
97202      Independent Movies          India     May 15, 2021
2021

       rating duration
description  \
146362  TV-MA  111 min  A dancer who gets a job cleaning at a
struggli...
31600   TV-MA  110 min  An American tourist goes on the run in
Greece ...
31621   TV-MA  115 min  Desperate to save their marriage, a young
coup...
97203   TV-MA   99 min  A struggling rickshaw driver’s life takes a
ro...
97202   TV-MA   81 min  Jilted by his wife, a man with OCD finds eye-
o...

                      cast
146362   Christopher Wollter
31600       Lena Kitsopoulou
31621           Diana Hopper
97203          Munivenkatapa
97202      Plabita Borthakur
```

Q6: *Show the top 5 TV Shows released based on their year of release*

```
tv_show_data=df1.loc[(df1["type"])=="TV Show"]
sorted_tv_show_data=tv_show_data.sort_values(["release_year"],ascendin
g=False)
sorted_tv_show_data.head()
```

```
       show_id     type                                  title  \
1           s9  TV Show         The Great British Baking Show
125664   s1294  TV Show                          Pitta Kathalu
121964    s199  TV Show  King of Boys: The Return of the King
121933    s110  TV Show                       La casa de papel
121931     s99  TV Show             Octonauts: Above & Beyond

               director                listed_in         country  \
1       Andy Devonshire         British TV Shows  United Kingdom
125664       Nag Ashwin                TV Dramas           India
121964     Kemi Adetiba   International TV Shows         Nigeria
121933        Mateo Gil   International TV Shows           Spain
121931    Jani Lachauer                Kids' TV  United Kingdom

                date_added  release_year rating   duration  \
1       September 24, 2021          2021  TV-14  9 Seasons
125664   February 19, 2021          2021  TV-MA   1 Season
```

```
121964     August 27, 2021          2021  TV-MA   1 Season
121933   September 3, 2021          2021  TV-MA  5 Seasons
121931   September 7, 2021          2021   TV-Y   1 Season

                                                  description  \
1       A talented batch of amateur bakers face off in...
125664  From secrets and lies to jealousy and control,...
121964  Alhaja Eniola Salami starts anew and sets her ...
121933  Eight thieves take hostages and lock themselve...
121931  The Octonauts expand their exploration beyond ...

                          cast
1                  Mel Giedroyc
125664                      Uko
121964   Keppy Ekpenyong Bassey
121933            Jaime Lorente
121931              Helen Walsh
```

MOVIES AND TV SHOWS IN INDIA

Q7: *Show only the titles of all the TV Shows that were released in India.*

```
df1[(df1["type"]=="TV Show") & (df1["country"]=="India")]["title"]
```

```
147                                 Navarasa
677                            Bombay Begums
717                            Pitta Kathalu
810       Mighty Little Bheem: Kite Festival
870                           Paava Kadhaigal
                          ...
150135                           Delhi Crime
152534                                   She
152789                           Delhi Crime
155005                           Delhi Crime
156648                           Delhi Crime
Name: title, Length: 940, dtype: object
```

Q8: *Show only the titles of all the Movies that were released in India.*

```
df1[(df1["type"]=="Movie") & (df1["country"]=="India")]["title"]
```

```
5                             Jeans
40                Angamaly Diaries
44                          Anjaam
46                          Dhanak
47                         Gurgaon
                    ...
159243        Haseena Maan Jaayegi
159246                      Pardes
159252      Y.M.I.: Yeh Mera India
159297        Haseena Maan Jaayegi
```

```
159300                    Pardes
Name: title, Length: 21150, dtype: object
```

Q8: *Show the list of Indian movie directors*

```
movie_directors=df1[(df1["type"]=="Movie") &
(df1["country"]=="India")]["director"]
```

```
movie_directors
```

```
5                  S. Shankar
40       Lijo Jose Pellissery
44                Rahul Rawail
46             Nagesh Kukunoor
47               Shanker Raman
                    ...
69556              David Dhawan
69566              Subhash Ghai
69582                N. Chandra
69730              David Dhawan
69739              Subhash Ghai
Name: director, Length: 7728, dtype: object
```

Q9: *Find the number of movies produced in each country and pick the top 10 countries.*

```
top_movies_country=df1[df1["type"]=="Movie"].groupby("country")
["title"].nunique().sort_values(ascending=False)
top_movies_country
```

```
country
United States     2143
India              939
 United States     346
United Kingdom     331
Canada             166
                   ...
Mozambique           1
Namibia              1
 Liechtenstein       1
 Latvia              1
Zimbabwe             1
Name: title, Length: 175, dtype: int64
```

Q10: *Find the number of Tv-Shows produced in each country and pick the top 10 countries.*

```
top_TVShows_country=df1[df1["type"]=="TV Show"].groupby("country")
["title"].nunique().sort_values(ascending=False)
top_TVShows_country
```

```
country
United States      325
United Kingdom     109
```

```
South Korea        107
Japan               95
Taiwan              40
Spain               27
India               24
Mexico              19
Turkey              16
France              15
 Canada             12
 United States      10
Thailand             9
Nigeria              7
Canada               6
Brazil               6
Colombia             5
Italy                5
Belgium              4
Poland               4
Denmark              4
Pakistan             3
Australia            2
 Germany             2
 Ireland             2
Philippines          2
 Spain               2
China                2
 France              2
Argentina            2
 South Africa        1
 Japan               1
 Italy               1
 Malaysia            1
 Mexico              1
 United Kingdom      1
Russia               1
Saudi Arabia         1
 Israel              1
 Indonesia           1
 South Korea         1
 Greece              1
 Singapore           1
Name: title, dtype: int64
```

Analysis of actors/directors of different movies and TV Shows

Q11: *The top 10 highest counts of "Movies" by each director.*

```
top_dir_movies=df1[df1["type"]=="Movie"]
top_dir_movies=top_dir_movies.drop_duplicates(["title","director"])
top_dir_movies["director"].value_counts().head(10)
```

```
Raúl Campos            20
Marcus Raboy           20
 Jan Suter             20
Rajiv Chilaka          16
Jay Karas              15
Cathy Garcia-Molina    13
Youssef Chahine        12
Martin Scorsese        12
Jay Chapman            12
Steven Spielberg       11
Name: director, dtype: int64
```

```
Q12: _The top 10 highest counts of “TV Shows” by each director_.
```

```
top_dir_TVShows=df1[df1["type"]=="TV Show"]
top_dir_TVShows=top_dir_TVShows.drop_duplicates(["title","director"])
top_dir_TVShows["director"].value_counts().head(10)
```

```
Hayato Date          70
Ryan Polito          61
Shin Won-ho          60
Michael Simon        52
Kenny Ortega         42
Hsu Fu-chun          31
Daniel Minahan       27
Jay Chandrasekhar    26
Ken Burns            25
Chang-Min Lee        20
Name: director, dtype: int64
```

BUSINESS INSIGHTS

- We got the top directors who has directed maximum number of movies and TV shows.

RECOMMENDATIONS

- We should add more movies and TV Shows of the above shows directors and they have received the maximum viewership

The top 10 actors who has acted in highest number of movies and TV Shows.

```
movies=df1[df1["type"]=="Movie"]
TV=df1[df1["type"]=="TV Show"]

movies=movies.drop_duplicates(["title","cast"])
TV=TV.drop_duplicates(["title","cast"])

movies["cast"].value_counts().head(5)
```

```
 Anupam Kher      38
 Om Puri          27
 Paresh Rawal     25
```

```
 Boman Irani      25
Shah Rukh Khan    25
Name: cast, dtype: int64
```

```
TV["cast"].value_counts().head(10)
```

```
 Takahiro Sakurai     15
 Ai Kayano            11
 Junichi Suwabe       11
 Katsuyuki Konishi     9
 Yuki Kaji             9
 Yoshimasa Hosoya      9
 Takehito Koyasu       9
 Daisuke Ono           8
 Kana Hanazawa         8
 Kohsuke Toriumi       8
Name: cast, dtype: int64
```

BUSINESS INSIGHTS

- We got the top actors who has acted in more number of movies and TV shows.

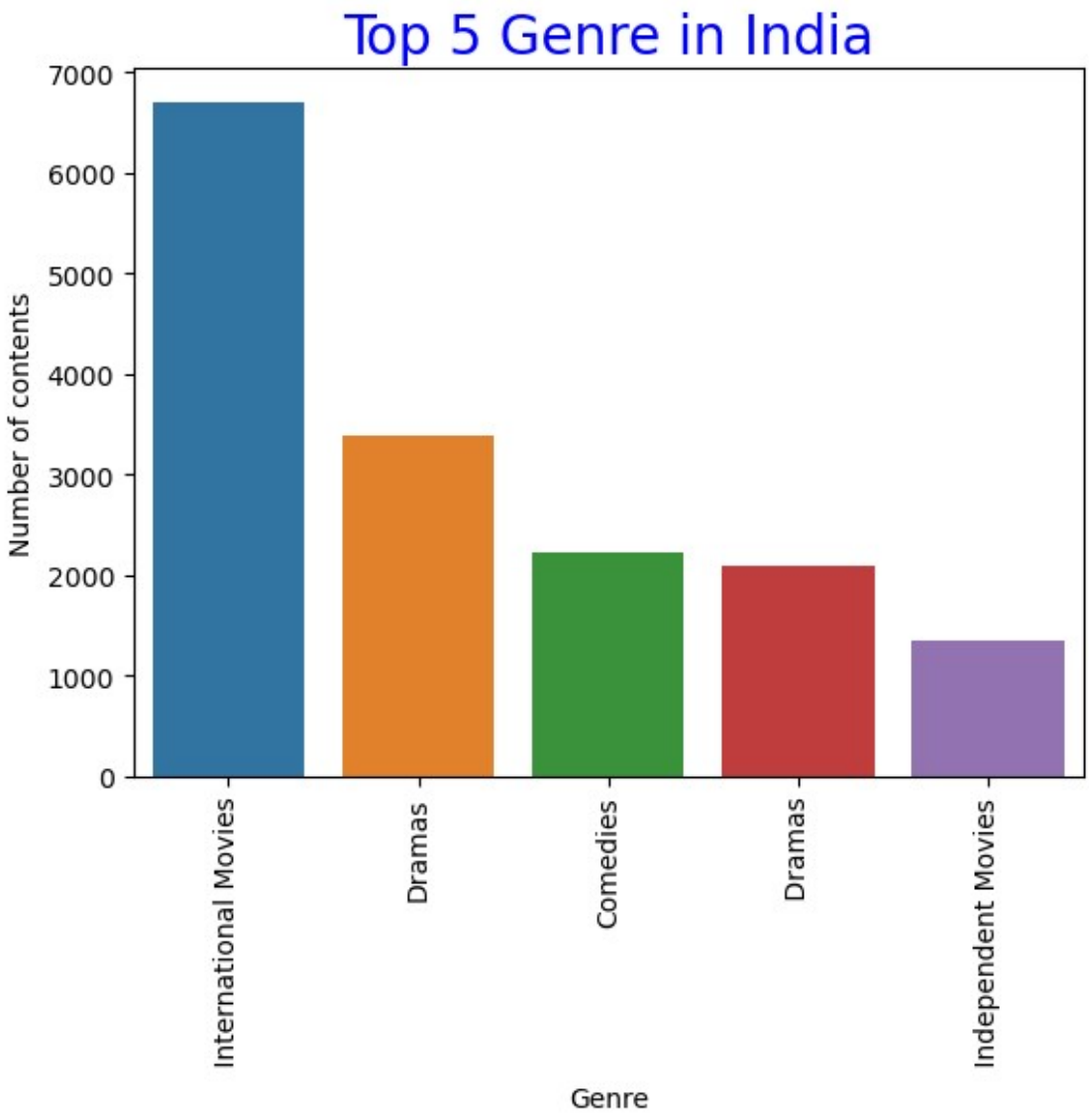
RECOMMENDATIONS

- We should add more movies and TV Shows of the above actors since they have received the maximum viewership

Q13: *Which genre of movies are popular in India,USA, and UK.*

```
ind=df1[df1["country"]=="India"]
genre_ind=ind["listed_in"].value_counts().head(5)
genre_ind=genre_ind.reset_index()
genre_ind.rename(columns={"index":"Genre", "listed_in": "Number of
contents"},inplace=True)
```

```
sns.barplot(data=genre_ind, x="Genre",y="Number of contents") #Boxplot
plt.xticks(rotation=90)
plt.title("Top 5 Genre in India",color= "BLUE",fontsize=20)
plt.show()
```

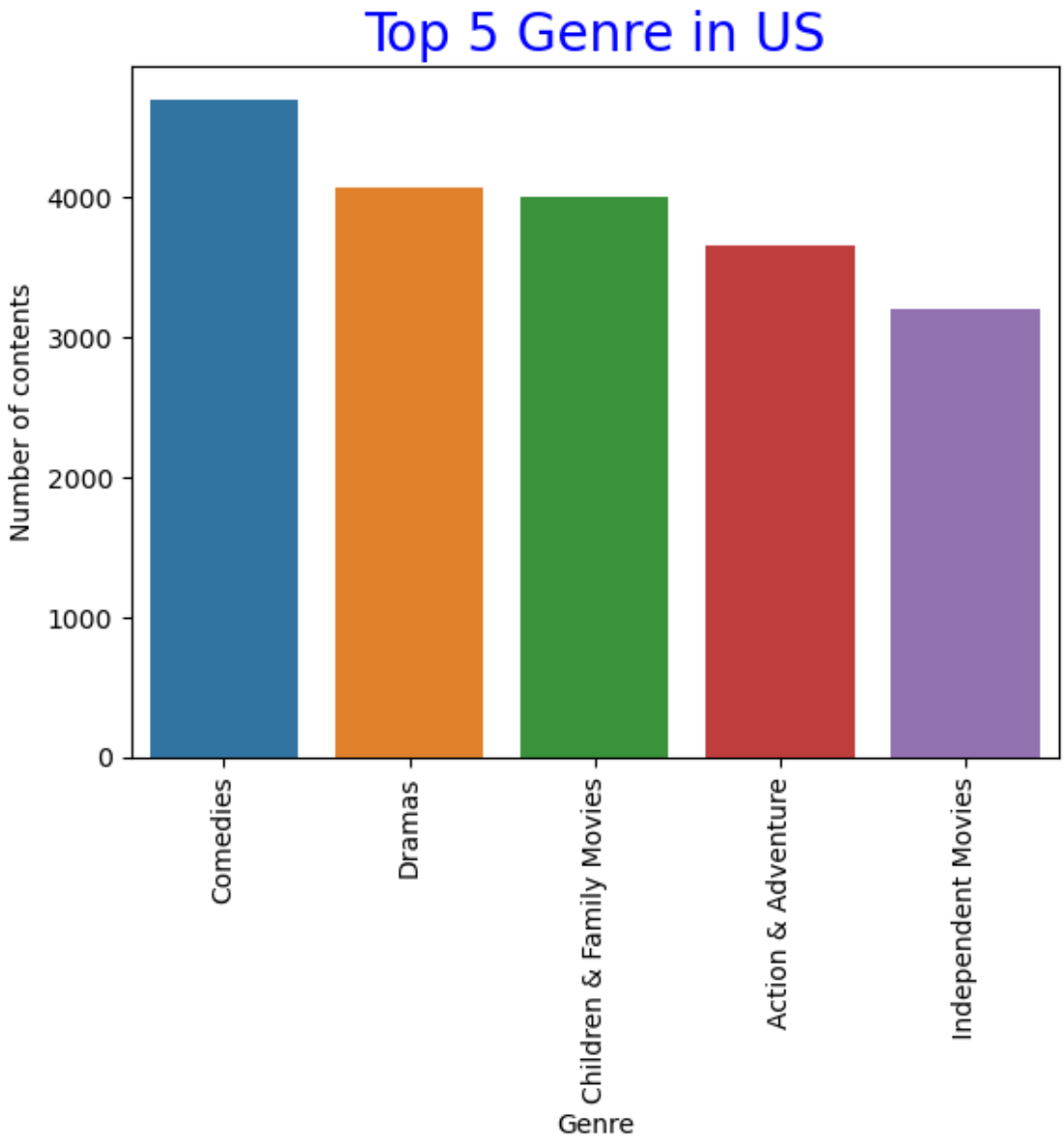


OBSERVATION

- International Movies and Dramas are the most popular genre in India

```
us=df1[df1["country"]=="United States"]
genre_us=us["listed_in"].value_counts().head(5)
genre_us=genre_us.reset_index()
genre_us.rename(columns={"index":"Genre", "listed_in": "Number of
contents"},inplace=True)

sns.barplot(data=genre_us, x="Genre",y="Number of contents") #Boxplot
plt.xticks(rotation=90)
plt.title("Top 5 Genre in US",color= "BLUE",fontsize=20)
plt.show()
```

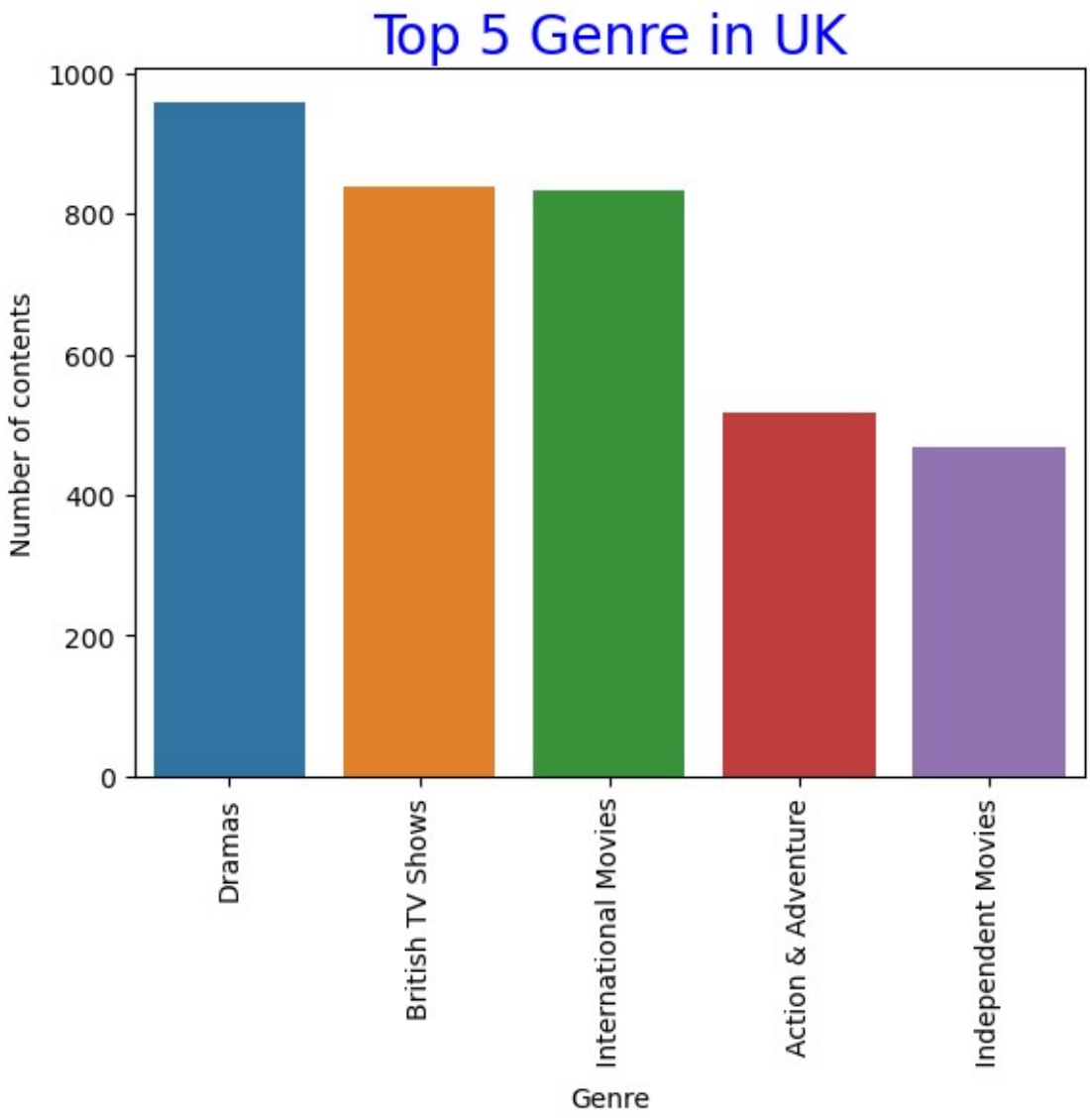


OBSERVATION

- Comedies and Dramas Shows are the most popular genre in the US

```
uk=df1[df1["country"]=="United Kingdom"]
genre_uk=uk["listed_in"].value_counts().head(5)
genre_uk=genre_uk.reset_index()
genre_uk.rename(columns={"index":"Genre", "listed_in": "Number of
contents"},inplace=True)

sns.barplot(data=genre_uk, x="Genre",y="Number of contents") #Boxplot
plt.xticks(rotation=90)
plt.title("Top 5 Genre in UK",color= "BLUE",fontsize=20)
plt.show()
```



OBSERVATION

- Dramas and British TV Shows are the most popular genre in the UK

BUSINESS INSIGHTS

- In India we can see that International movies & Dramas are the most popular genres.
- In the United States Comedies & Dramas are most popular genre while, independent movies genre is the least popular.
- In the United Kingdom audience is more into Dramas.

RECOMMENDATIONS

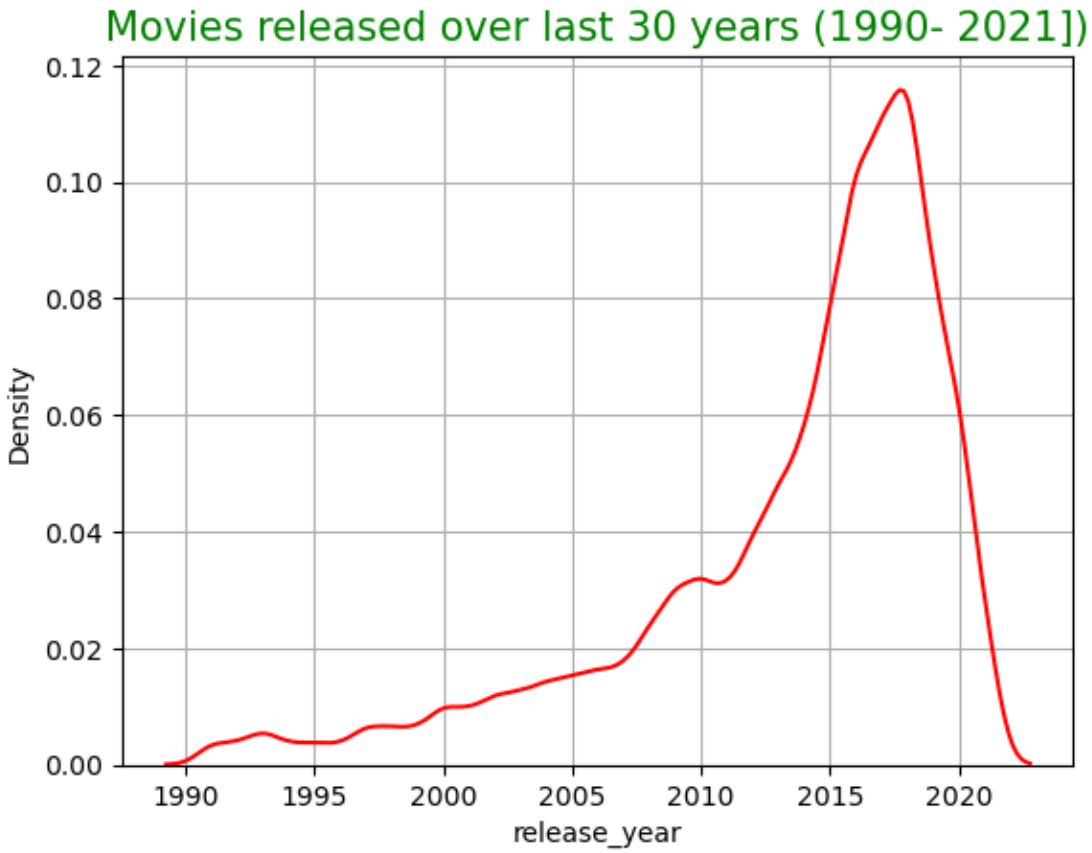
- We should add more contents related to genre like International movies & Dramas in India as Indians are watching more of it.

- For United States, we should add more comedies & Dramas genre content as compare to other genre.
- In United Kingdom, similar as US, we should add more content related to genre like Dramas.

Q14: *How has the number of movies released per year changed over last 20-30 years?.*

```
last30=df2[(df2["release_year"]>1990) & (df2["type"]=="Movie")]
```

```
sns.kdeplot(data = last30, x = "release_year", color="red")
plt.title("Movies released over last 30 years (1990-
2021])",color="green",fontsize=15)
plt.grid()
plt.show()
```



OBSERVATION

- After analyzing the graph of last 30 years, we can say that from 2011 to 2018 number of movies released in a huge number. While, we can also notice that there is a sudden decrement in release after 2018-19 and that could be due to the Covid-19 disease.

Q14: *What is the best time to launch TV Shows?.*

```
df2 = df1
df2['month'] = pd.to_datetime(df2['date_added']).dt.month_name()

df2.head(5)
```

```
  show_id     type                          title             director
\
0      s8    Movie                         Sankofa         Haile Gerima

1      s9  TV Show  The Great British Baking Show      Andy Devonshire

2     s10    Movie                    The Starling       Theodore Melfi

3     s13    Movie                    Je Suis Karl  Christian Schwochow

4     s16  TV Show               Dear White People    Jay Chandrasekhar


          listed_in         country          date_added  release_year
rating  \
0           Dramas   United States  September 24, 2021          1993
TV-MA
1  British TV Shows  United Kingdom  September 24, 2021          2021
TV-14
2          Comedies   United States  September 24, 2021          2021
PG-13
3           Dramas         Germany  September 23, 2021          2021
TV-MA
4       TV Comedies   United States  September 22, 2021          2021
TV-MA

    duration                                           description  \
0    125 min  On a photo shoot in Ghana, an American model s...
1  9 Seasons  A talented batch of amateur bakers face off in...
2    104 min  A woman adjusting to life after a loss contend...
3    127 min  After most of her family is murdered in a terr...
4  4 Seasons  Students of color navigate the daily slights a...

                cast      month
0      Kofi Ghanaba  September
1      Mel Giedroyc  September
2  Melissa McCarthy  September
3       Luna Wedler  September
4    Logan Browning  September
```

```
tv_show_counts = df2[df2["type"] == 'TV Show'].groupby(["month"])
["title"].count()

tv_show_counts
```

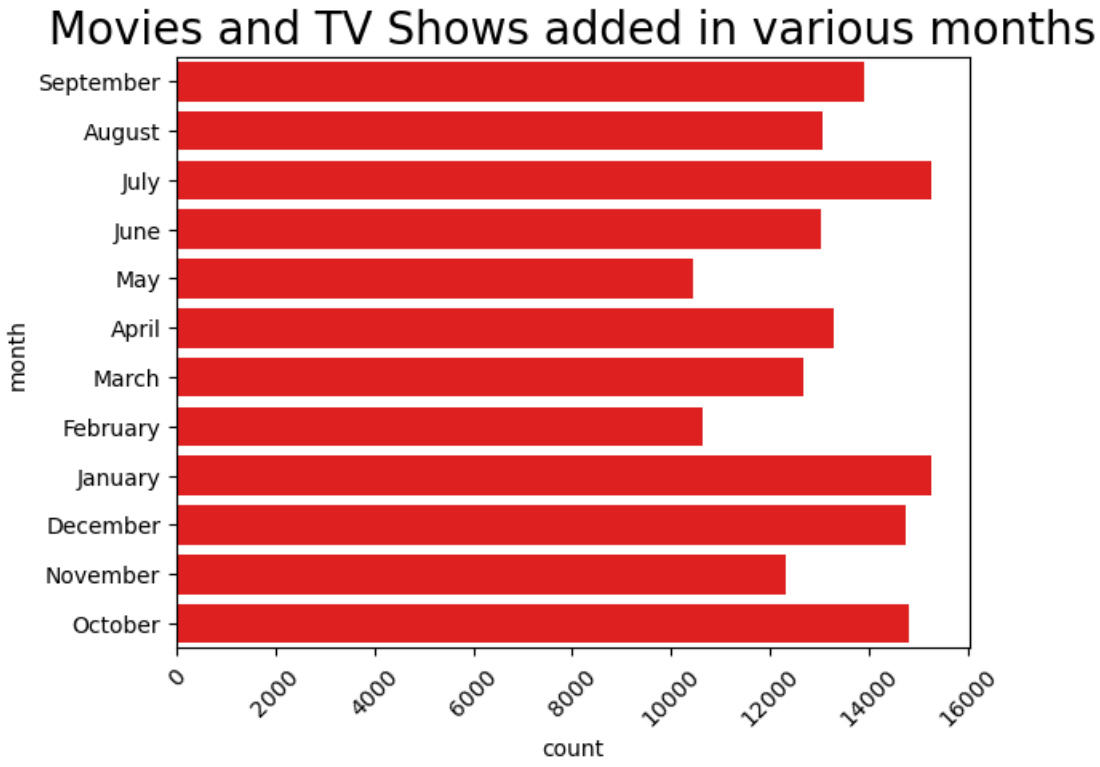
```
month
1     1691
2     1815
3     1581
4     1328
5     1614
6     2343
7     1486
8     1933
9     1392
10    1576
11    1477
12    2491
Name: title, dtype: int64
```

```
Movie_counts = df2[df2["type"] == 'Movie'].groupby(["month"])
["title"].count()
Movie_counts
```

```
month
1     13584
2      8837
3     11110
4     11949
5      8836
6     10697
7     13789
8     11145
9     12514
10    13235
11    10837
12    12267
Name: title, dtype: int64
```

```
data=df2['month']
data=data.reset_index()
```

```
sns.countplot(data=data,y=df2['month'],color="red")
plt.title("Movies and TV Shows added in various months",fontsize=20)
plt.xticks(rotation=45)
plt.show()
```



BUSINESS INSIGHTS

-July and January are the best time to launch a TV show or movies, after that October,December can be considered as best time.

RECOMMENDATIONS

-We should release majorly of the content during July, December, October and January month.

Rating of Movies and TV Shows in United States

```
movies_us=df1[(df1["type"]=="Movie") & (df1["country"]=="United States")]
tv_us=df1[(df1["type"]=="TV Show") & (df1["country"]=="United States")]


movies_us=movies_us.drop_duplicates(["title"])
tv_us=tv_us.drop_duplicates(["title"])

movies_us["rating"].value_counts().head(5)
```

```
TV-MA    560
R        529
PG-13    354
PG       207
```

```
TV-14    177
Name: rating, dtype: int64
```

```
tv_us["rating"].value_counts().head(5)
```

```
TV-MA    136
TV-14     87
TV-PG     51
TV-G      23
TV-Y7     18
Name: rating, dtype: int64
```

BUSINESS INSIGHTS

-We have extracted the top rating category who have the maximum number of movies in the US market.

RECOMMENDATIONS

-We should add more movies related to above mentioned rating category as it will going to add more viewership on our portal.

TOP 10 POPULAR TV SHOWS

```
TV=df1[df1["type"]=="TV Show"]
TV=TV.drop_duplicates(["title"])
TV[["title","duration"]].sort_values(by="duration",ascending=False).he
ad(10).reset_index(drop=True)
```

	title	duration
0	The Great British Baking Show	9 Seasons
1	Saved by the Bell	9 Seasons
2	Call the Midwife	9 Seasons
3	Naruto	9 Seasons
4	The Office (U.S.)	9 Seasons
5	American Horror Story	9 Seasons
6	Trolls: The Beat Goes On!	8 Seasons
7	Dexter	8 Seasons
8	Girlfriends	8 Seasons
9	Arrow	8 Seasons

BUSINESS INSIGHTS

-We have extracted the top 10 popular TV Shows

RECOMMENDATIONS

-As these shows are huge hit ,we should make their next seasons as well due to their high popularity and viewership

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

```
df2 = df1
df2['Year'] = pd.to_datetime(df2['date_added']).dt.year

df3= df3[(df3["Year"] >= 2015) & (df2["Year"] <= 2021)]
recent_TVshow_count = df3[df3["type"] == 'TV Show'].groupby("Year")
["title"].count()
recent_Movie_count = df3[df3["type"] == 'Movie'].groupby("Year")
["title"].count()

recent_TVshow_count
```

```
Year
2015     172
2016    1401
2017    2917
2018    2629
2019    4901
2020    5548
2021    3103
Name: title, dtype: int64
```

```
recent_Movie_count
```

```
Year
2015     1062
2016     4687
2017    17838
2018    27551
2019    33284
2020    31598
2021    21827
Name: title, dtype: int64
```

***** THANK YOU *****