```
!unzip netflix_titles.csv.zip
```

```
Archive:  netflix_titles.csv.zip
  inflating: netflix_titles.csv
```

```
#importing libraries for our purpose
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv('netflix_titles.csv')
df.head()
```

	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...
					Sami						Crime TV	To protect his

```
#length of data
len(df)
```

```
8807
```

```
#checking datatypes
df.dtypes
```

```
show_id         object
type            object
title           object
director        object
cast            object
country         object
date_added      object
release_year     int64
rating          object
duration        object
listed_in       object
description     object
dtype: object
```

```
#number of unique values in our data
for i in df.columns:
  print(i,':',df[i].nunique())
```

```
show_id : 8807
type : 2
title : 8807
director : 4528
cast : 7692
country : 748
date_added : 1767
release_year : 74
rating : 17
duration : 220
listed_in : 514
description : 8775
```

```
#checking null values in every column of our data
df.isnull().sum()
```

```
show_id            0
type               0
title              0
director        2634
cast             825
country          831
date_added        10
release_year       0
rating             4
duration           3
listed_in          0
description        0
dtype: int64
```

```
#checking the occurences of each of the ratings
df['rating'].value_counts()
```

```
TV-MA       3207
TV-14       2160
TV-PG        863
R            799
PG-13        490
TV-Y7        334
TV-Y         307
PG           287
TV-G         220
NR            80
G             41
TV-Y7-FV       6
NC-17          3
UR             3
74 min         1
84 min         1
66 min         1
Name: rating, dtype: int64
```

```
con1 = df['director'].apply(lambda x: str(x).split(', ')).tolist()
```

```
df_new1=pd.DataFrame(con1,index=df['title'])
df_new1
```

title	0	1	2	3	4	5	6	7	8	9	10	11	12
Dick Johnson Is Dead	Kirsten Johnson	None	None	None	None	None	None	None	None	None	None	None	None
Blood & Water	nan	None	None	None	None	None	None	None	None	None	None	None	None
Ganglands	Julien Leclercq	None	None	None	None	None	None	None	None	None	None	None	None
Jailbirds New Orleans	nan	None	None	None	None	None	None	None	None	None	None	None	None
Kota Factory	nan	None	None	None	None	None	None	None	None	None	None	None	None
...	...	...	...	...	...	...	...	...	...	...	...	...	...
Zodiac	David Fincher	None	None	None	None	None	None	None	None	None	None	None	None
Zombie Dumb	nan	None	None	None	None	None	None	None	None	None	None	None	None
Zombieland	Ruben Fleischer	None	None	None	None	None	None	None	None	None	None	None	None
Zoom	Peter Hewitt	None	None	None	None	None	None	None	None	None	None	None	None
Zubaan	Mozez Singh	None	None	None	None	None	None	None	None	None	None	None	None

8807 rows × 13 columns

```
#unnesting the directors column, i.e- creating separate lines for each director in a movie
constraint1=df['director'].apply(lambda x: str(x).split(', ')).tolist()
df_new1=pd.DataFrame(constraint1,index=df['title'])
df_new1=df_new1.stack()
df_new1=pd.DataFrame(df_new1.reset_index())
df_new1.rename(columns={0:'Directors'},inplace=True)
df_new1.drop(['level_1'],axis=1,inplace=True)
df_new1.head()
```

	title	Directors
0	Dick Johnson Is Dead	Kirsten Johnson
1	Blood & Water	nan
2	Ganglands	Julien Leclercq
3	Jailbirds New Orleans	nan
4	Kota Factory	nan

```
#unnesting the cast column, i.e- creating separate lines for each cast member in a movie
constraint2=df['cast'].apply(lambda x: str(x).split(', ')).tolist()
df_new2=pd.DataFrame(constraint2,index=df['title'])
df_new2=df_new2.stack()
df_new2=pd.DataFrame(df_new2.reset_index())
df_new2.rename(columns={0:'Actors'},inplace=True)
df_new2.drop(['level_1'],axis=1,inplace=True)
df_new2.head()
```

	title	Actors
0	Dick Johnson Is Dead	nan
1	Blood & Water	Ama Qamata
2	Blood & Water	Khosi Ngema
3	Blood & Water	Gail Mabalane
4	Blood & Water	Thabang Molaba

```
#unnesting the listed_in column, i.e- creating separate lines for each genre in a movie
constraint3=df['listed_in'].apply(lambda x: str(x).split(', ')).tolist()
df_new3=pd.DataFrame(constraint3,index=df['title'])
df_new3=df_new3.stack()
df_new3=pd.DataFrame(df_new3.reset_index())
df_new3.rename(columns={0:'Genre'},inplace=True)
df_new3.drop(['level_1'],axis=1,inplace=True)
df_new3.head()
```

	title	Genre
0	Dick Johnson Is Dead	Documentaries
1	Blood & Water	International TV Shows
2	Blood & Water	TV Dramas
3	Blood & Water	TV Mysteries
4	Ganglands	Crime TV Shows

```
#unnesting the country column, i.e- creating separate lines for each country in a movie
constraint4=df['country'].apply(lambda x: str(x).split(', ')).tolist()
df_new4=pd.DataFrame(constraint4,index=df['title'])
df_new4=df_new4.stack()
df_new4=pd.DataFrame(df_new4.reset_index())
df_new4.rename(columns={0:'country'},inplace=True)
df_new4.drop(['level_1'],axis=1,inplace=True)
df_new4.head()
```

	title	country
0	Dick Johnson Is Dead	United States
1	Blood & Water	South Africa
2	Ganglands	nan
3	Jailbirds New Orleans	nan
4	Kota Factory	India

```
#merging the unnested director data with unnested actors data
dfx
```

	title	Actors	Directors	Genre	country
0	Dick Johnson Is Dead	Unknown Actor	Kirsten Johnson	Documentaries	United States
1	Blood & Water	Ama Qamata	Unknown Director	International TV Shows	South Africa
2	Blood & Water	Ama Qamata	Unknown Director	TV Dramas	South Africa
3	Blood & Water	Ama Qamata	Unknown Director	TV Mysteries	South Africa
4	Blood & Water	Khosi Ngema	Unknown Director	International TV Shows	South Africa

```
#merging our unnested data with the original data
df_final=df_new.merge(df[['show_id', 'type', 'title', 'date_added',
       'release_year', 'rating', 'duration']],on=['title'],how='left')
df_final.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration
0	Dick Johnson Is Dead	Unknown Actor	Kirsten Johnson	Documentaries	United States	s1	Movie	September 25, 2021	2020	PG-13	90 min
1	Blood & Water	Ama Qamata	Unknown Director	International TV Shows	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons
2	Blood & Water	Ama Qamata	Unknown Director	TV Dramas	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons

```
#now checking nulls
df_final.isnull().sum()
```

```
title              0
Actors             0
Directors          0
Genre              0
country        11897
show_id            0
type               0
date_added       158
release_year       0
rating            67
duration           3
dtype: int64
```

In duration column, it was observed that the nulls had values which were written in corresponding ratings column, i.e- you can't expect ratings to be in min. So the duration column nulls are replaced by corresponding values in ratings column

```
df_final.loc[df_final['duration'].isnull(),'duration']=df_final.loc[df_final['duration'].isnull(),'duration'].fillna(df_final['rating'])
df_final.loc[df_final['rating'].str.contains('min', na=False),'rating']='NR'
df_final.isnull().sum()
```

```
title              0
Actors             0
Directors          0
Genre              0
country        11897
show_id            0
type               0
date_added       158
release_year       0
rating            67
duration           0
dtype: int64
```

```
#Ratings can't be in min, so it has been made NR(i.e- Non Rated)
df_final.loc[df_final['rating'].str.contains('min', na=False),'rating']='NR'
df_final['rating'].fillna('NR',inplace=True)
pd.set_option('display.max_rows',None)


#just an attempt to observe nulls in date_added column
df_final[df_final['date_added'].isnull()].head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration
136893	A Young Doctor's Notebook and Other Stories	Daniel Radcliffe	Unknown Director	British TV Shows	United Kingdom	s6067	TV Show	NaN	2013	TV-MA	2 Seasons
136894	A Young Doctor's Notebook and Other Stories	Daniel Radcliffe	Unknown Director	TV Comedies	United Kingdom	s6067	TV Show	NaN	2013	TV-MA	2 Seasons

```
#date added column is imputed on the basis of release year,i.e- suppose there's a null for date_added
#when release year was 2013.So below piece of code just checks the mode of date added for release year=2013
# and imputes in place of nulls the corresponding mode

for i in df_final[df_final['date_added'].isnull()]['release_year'].unique():
  imp=df_final[df_final['release_year']==i]['date_added'].mode().values[0]
  df_final.loc[df_final['release_year']==i,'date_added']=df_final.loc[df_final['release_year']==i,'date_added'].fillna(imp)


#country column is imputed on the basis of director,i.e- suppose there's a null for country
#when we have a director whose other movies have a country given.So below piece of code just checks the mode of
#country for the director
# and imputes in place of nulls the corresponding mode

for i in df_final[df_final['country'].isnull()]['Directors'].unique():
  if i in df_final[~df_final['country'].isnull()]['Directors'].unique():
    imp=df_final[df_final['Directors']==i]['country'].mode().values[0]
    df_final.loc[df_final['Directors']==i,'country']=df_final.loc[df_final['Directors']==i,'country'].fillna(imp)
```

So we imputed the country column on the basis of directors whose other movie titles had countries given. But there might be directors who have only one occurence in our data. In that scenario, I have used Actors as a basis. i.e- for this Actor majorly acts in movies of which country? Imputation has been done on this basis. For remaining rows, country has been filled as Unknown Country

```
for i in df_final[df_final['country'].isnull()]['Actors'].unique():
  if i in df_final[~df_final['country'].isnull()]['Actors'].unique():
    imp=df_final[df_final['Actors']==i]['country'].mode().values[0]
    df_final.loc[df_final['Actors']==i,'country']=df_final.loc[df_final['Actors']==i,'country'].fillna(imp)
#If there are still nulls, I just replace it by Unknown Country
df_final['country'].fillna('Unknown Country',inplace=True)
df_final.isnull().sum()
```

```
title           0
Actors          0
Directors       0
Genre           0
country         0
show_id         0
type            0
date_added      0
release_year    0
rating          0
duration        0
dtype: int64
```

```
df_final.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration
0	Dick Johnson Is Dead	Unknown Actor	Kirsten Johnson	Documentaries	United States	s1	Movie	September 25, 2021	2020	PG-13	90 min
1	Blood & Water	Ama Qamata	Unknown Director	International TV Shows	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons
2	Blood & Water	Ama Qamata	Unknown Director	TV Dramas	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons

```
df_final['duration'].value_counts()
```

```
1 Season      35035
2 Seasons      9559
3 Seasons      5084
94 min         4343
106 min        4040
97 min         3624
95 min         3560
96 min         3484
93 min         3480
90 min         3305
```

```
105 min        3209
107 min        3103
101 min        3048
102 min        3017
103 min        2985
98 min         2984
99 min         2956
91 min         2915
92 min         2863
104 min        2822
88 min         2781
110 min        2711
100 min        2697
108 min        2614
112 min        2594
85 min         2486
89 min         2420
86 min         2213
4 Seasons      2134
116 min        2122
118 min        2119
119 min        2075
87 min         2063
109 min        2020
113 min        1990
120 min        1845
117 min        1770
121 min        1728
5 Seasons      1698
111 min        1667
124 min        1590
114 min        1529
127 min        1505
115 min        1444
123 min        1398
125 min        1299
122 min        1298
84 min         1268
128 min        1241
130 min        1216
126 min        1205
81 min         1203
83 min         1192
133 min        1169
137 min        1122
82 min         1100
136 min        1092
132 min        1047
```

```
#removing mins from data
df_final['duration']=df_final['duration'].str.replace(" min","")
df_final.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration
0	Dick Johnson Is Dead	Unknown Actor	Kirsten Johnson	Documentaries	United States	s1	Movie	September 25, 2021	2020	PG-13	90
1	Blood & Water	Ama Qamata	Unknown Director	International TV Shows	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons
2	Blood & Water	Ama Qamata	Unknown Director	TV Dramas	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons

```
df_final['duration'].unique()
```

```
array(['90', '2 Seasons', '1 Season', '91', '125', '9 Seasons', '104',
       '127', '4 Seasons', '67', '94', '5 Seasons', '161', '61', '166',
       '147', '103', '97', '106', '111', '3 Seasons', '110', '105', '96',
       '124', '116', '98', '23', '115', '122', '99', '88', '100',
       '6 Seasons', '102', '93', '95', '85', '83', '113', '13', '182',
       '48', '145', '87', '92', '80', '117', '128', '119', '143', '114',
       '118', '108', '63', '121', '142', '154', '120', '82', '109', '101',
       '86', '229', '76', '89', '156', '112', '107', '129', '135', '136',
```

```
       '165', '150', '133', '70', '84', '140', '78', '7 Seasons', '64',
       '59', '139', '69', '148', '189', '141', '130', '138', '81', '132',
       '10 Seasons', '123', '65', '68', '66', '62', '74', '131', '39',
       '46', '38', '8 Seasons', '17 Seasons', '126', '155', '159', '137',
       '12', '273', '36', '34', '77', '60', '49', '58', '72', '204',
       '212', '25', '73', '29', '47', '32', '35', '71', '149', '33', '15',
       '54', '224', '162', '37', '75', '79', '55', '158', '164', '173',
       '181', '185', '21', '24', '51', '151', '42', '22', '134', '177',
       '13 Seasons', '52', '14', '53', '8', '57', '28', '50', '9', '26',
       '45', '171', '27', '44', '146', '20', '157', '17', '203', '41',
       '30', '194', '15 Seasons', '233', '237', '230', '195', '253',
       '152', '190', '160', '208', '180', '144', '5', '174', '170', '192',
       '209', '187', '172', '16', '186', '11', '193', '176', '56', '169',
       '40', '10', '3', '168', '312', '153', '214', '31', '163', '19',
       '12 Seasons', '179', '11 Seasons', '43', '200', '196', '167',
       '178', '228', '18', '205', '201', '191'], dtype=object)
```

```
df_final['duration_copy']=df_final['duration'].copy()
df_final1=df_final.copy()
```

```
df_final1.loc[df_final1['duration_copy'].str.contains('Season'),'duration_copy']=0
df_final1['duration_copy']=df_final1['duration_copy'].astype('int')
df_final1.head()
```

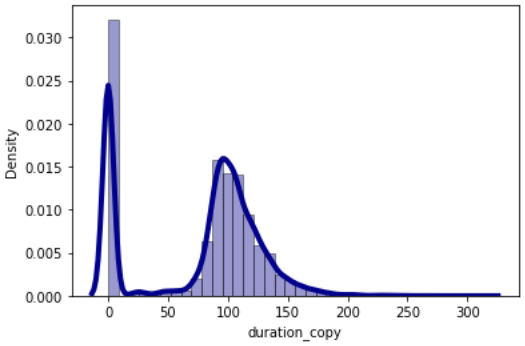
	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	duration_copy
0	Dick Johnson Is Dead	Unknown Actor	Kirsten Johnson	Documentaries	United States	s1	Movie	September 25, 2021	2020	PG-13	90	90
1	Blood & Water	Ama Qamata	Unknown Director	International TV Shows	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	0
2	Blood &	Ama	Unknown	TV Dramas	South	s2	TV	September	2021	TV-MA	2	0

```
df_final1['duration_copy'].describe()
```

```
count    201991.000000
mean         77.152789
std          52.269154
min           0.000000
25%           0.000000
50%          95.000000
75%         112.000000
max         312.000000
Name: duration_copy, dtype: float64
```

```
import seaborn as sns
sns.distplot(df_final1['duration_copy'], hist=True, kde=True,
bins=int(36), color = 'darkblue',
hist_kws={'edgecolor':'black'},
kde_kws={'linewidth': 4})
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be rem
  warnings.warn(msg, FutureWarning)
```




```
bins1 = [-1,1,50,80,100,120,150,200,315]
labels1 = ['<1','1-50','50-80','80-100','100-120','120-150','150-200','200-315']
df_final1['duration_copy'] = pd.cut(df_final1['duration_copy'],bins=bins1,labels=labels1)
df_final1.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	duration_copy
0	Dick Johnson Is Dead	Unknown Actor	Kirsten Johnson	Documentaries	United States	s1	Movie	September 25, 2021	2020	PG-13	90	80-100
1	Blood & Water	Ama Qamata	Unknown Director	International TV Shows	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	<1
2	Blood &	Ama	Unknown	TV Dramas	South	s2	TV	September	2021	TV-MA	2	<1

```
df_final1.loc[~df_final1['duration'].str.contains('Season'),'duration']=df_final1.loc[~df_final1['duration'].str.contains('Season'),'duratic
df_final1.drop(['duration_copy'],axis=1,inplace=True)
df_final1.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration
0	Dick Johnson Is Dead	Unknown Actor	Kirsten Johnson	Documentaries	United States	s1	Movie	September 25, 2021	2020	PG-13	80-100
1	Blood & Water	Ama Qamata	Unknown Director	International TV Shows	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons
2	Blood & Water	Ama Qamata	Unknown Director	TV Dramas	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons

```
df_final1['duration'].value_counts()
```

```
80-100        52937
100-120       48724
1 Season      35035
120-150       26691
2 Seasons      9559
50-80          7700
150-200        6737
3 Seasons      5084
1-50           2530
4 Seasons      2134
5 Seasons      1698
```

```
7 Seasons       843
6 Seasons       633
200-315         524
8 Seasons       286
9 Seasons       257
10 Seasons      220
13 Seasons      132
12 Seasons      111
15 Seasons       96
17 Seasons       30
11 Seasons       30
Name: duration, dtype: int64
```

```
from datetime import datetime
from dateutil.parser import parse
arr=[]
for i in df_final1['date_added'].values:
  dt1=parse(i)
  arr.append(dt1.strftime('%Y-%m-%d'))
df_final1['Modified_Added_date'] =arr
df_final1['Modified_Added_date']=pd.to_datetime(df_final1['Modified_Added_date'])
df_final1['month_added']=df_final1['Modified_Added_date'].dt.month
df_final1['week_Added']=df_final1['Modified_Added_date'].dt.week
df_final1['year']=df_final1['Modified_Added_date'].dt.year
df_final1.head()
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:10: FutureWarning: Series.dt.weekofyear and Series.dt.week have been deprec
  # Remove the CWD from sys.path while we load stuff.
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	Modified_Added_date	r
0	Dick Johnson Is Dead	Unknown Actor	Kirsten Johnson	Documentaries	United States	s1	Movie	September 25, 2021	2020	PG-13	80-100	2021-09-25	
1	Blood & Water	Ama Qamata	Unknown Director	International TV Shows	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	2021-09-24	
2	Blood & Water	Ama Qamata	Unknown Director	TV Dramas	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	2021-09-24	
3	Blood & Water	Ama Qamata	Unknown Director	TV Mysteries	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	2021-09-24	
4	Blood & Water	Khosi Ngema	Unknown Director	International TV Shows	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	2021-09-24	

```
#Titles such as Bahubali(Hindi Version),Bahubali(Tamil Version) were there. Since it's only one movie in different languages,
#presence of brackets and content between brackets is removed.
df_final1['title']=df_final1['title'].str.replace(r"\(.*\)","")
df_final1.head()
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: FutureWarning: The default value of regex will change from True to False
  This is separate from the ipykernel package so we can avoid doing imports until
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	Modified_Added_date	r
0	Dick Johnson Is Dead	Unknown Actor	Kirsten Johnson	Documentaries	United States	s1	Movie	September 25, 2021	2020	PG-13	80-100	2021-09-25	
1	Blood & Water	Ama Qamata	Unknown Director	International TV Shows	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	2021-09-24	
2	Blood & Water	Ama Qamata	Unknown Director	TV Dramas	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	2021-09-24	
3	Blood & Water	Ama Qamata	Unknown Director	TV Mysteries	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	2021-09-24	
	Blood &	Khosi	Unknown	International	South		TV	September			2		

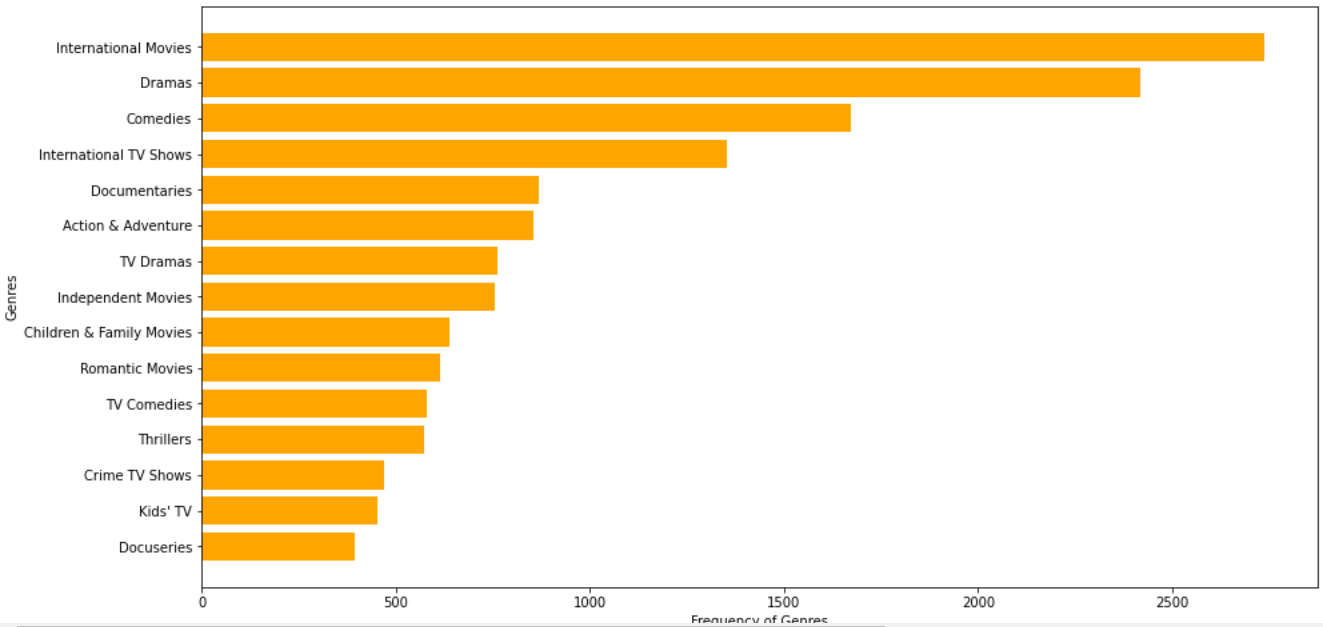
Univariate Analysis in terms of counts of each column

```
#number of distinct titles on the basis of genre
df_final1.groupby(['Genre']).agg({"title":"nunique"})
```

Genre	title
Action & Adventure	854
Anime Features	71
Anime Series	176
British TV Shows	253
Children & Family Movies	639
Classic & Cult TV	28
Classic Movies	116
Comedies	1673
Crime TV Shows	470
Cult Movies	71
Documentaries	869
Docuseries	395
Dramas	2418
Faith & Spirituality	65
Horror Movies	353
Independent Movies	756
International Movies	2738
International TV Shows	1351
Kids' TV	451
Korean TV Shows	151
LGBTQ Movies	102
Movies	57
Music & Musicals	372
Reality TV	255
Romantic Movies	615
Romantic TV Shows	370
Sci-Fi & Fantasy	243
Science & Nature TV	92
Spanish-Language TV Shows	174
Sports Movies	219
Stand-Up Comedy	343
Stand-Up Comedy & Talk Shows	56
TV Action & Adventure	168
TV Comedies	581
TV Dramas	763
TV Horror	75
TV Mysteries	98
TV Sci-Fi & Fantasy	84
TV Shows	16
TV Thrillers	57
Teen TV Shows	69
Thrillers	573

```
df_genre=df_final1.groupby(['Genre']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_genre[::-1]['Genre'], df_genre[::-1]['title'],color=['orange'])
```

```
plt.xlabel('Frequency of Genres')
plt.ylabel('Genres')
plt.show()
```

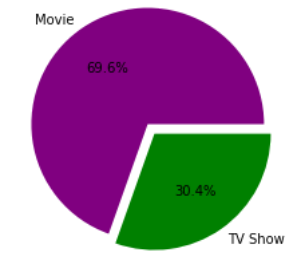


International Movies, Dramas and Comedies are the most popular .

```
#number of distinct titles on the basis of type
df_final1.groupby(['type']).agg({"title":"nunique"})
```

type	title
Movie	6115
TV Show	2676

```
df_type=df_final1.groupby(['type']).agg({"title":"nunique"}).reset_index()
plt.pie(df_type['title'],explode=(0.05,0.05), labels=df_type['type'],colors=['purple','green'],autopct='%.1f%%')
plt.show()
```



We have 70:30 ratio of Movies and TV Shows in our data

```
#number of distinct titles on the basis of country
df_final1.groupby(['country']).agg({"title":"nunique"})
```

country	title
	3
Afghanistan	1
Albania	1
Algeria	3
Angola	2
Argentina	94
Armenia	1
Australia	162
Austria	12
Azerbaijan	1
Bahamas	1
Bangladesh	4
Belarus	1
Belgium	94
Bermuda	1
Botswana	1
Brazil	103
Bulgaria	10
Burkina Faso	1
Cambodia	5
Cambodia,	1
Cameroon	2
Canada	460
Cayman Islands	2
Chile	30
China	166
Colombia	54
Croatia	4
Cuba	2
Cyprus	1
Czech Republic	23
Denmark	50
Dominican Republic	1
East Germany	1
Ecuador	1
Egypt	134
Ethiopia	1
Finland	12
France	409
Georgia	2
Germany	231
Ghana	8
Greece	11
Guatemala	2
Hong Kong	110

Hungary	11
Iceland	11
India	1126
Indonesia	97
Iran	4
Iraq	2
Ireland	46
Israel	30
Italy	102
Jamaica	1
Japan	338
Jordan	10
Kazakhstan	1
Kenya	6
Kuwait	9
Latvia	1
Lebanon	33
Liechtenstein	1
Lithuania	1
Luxembourg	12
Malawi	1
Malaysia	26
Malta	3
Mauritius	3
Mexico	175
Mongolia	1
Montenegro	1
Morocco	6
Mozambique	1
Namibia	2
Nepal	2
Netherlands	50
New Zealand	33
Nicaragua	1
Nigeria	140
Norway	30
Pakistan	24
Palestine	1
Panama	1
Paraguay	1
Peru	11
Philippines	90
Poland	41
Poland,	1
Portugal	6
Puerto Rico	1
Qatar	10

Qatar	10
Romania	14
Russia	27
Samoa	1
Saudi Arabia	14
Senegal	3
Serbia	7
Singapore	41
Slovakia	1
Slovenia	3
Somalia	1
South Africa	65
South Korea	235
Soviet Union	3
Spain	239
Sri Lanka	1
Sudan	1
Sweden	44
Switzerland	19
Syria	3
Taiwan	94
Thailand	74
Turkey	115
Uganda	1
Ukraine	3
United Arab Emirates	38
United Kingdom	829
United Kingdom,	2
United States	4245
United States,	1
Unknown Country	175
Uruguay	14
Vatican City	1
Venezuela	4
Vietnam	7
West Germany	5
Zimbabwe	3

The above dataframe shows a flaw in which we are seeing countries, such as Cambodia and Cambodia, or United States and United States, are shown as different countries.They should have been same

```
df_final1['country'] = df_final1['country'].str.replace(',', '')
df_final1.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	Modified_Added_date	r
0	Dick Johnson Is Dead	Unknown Actor	Kirsten Johnson	Documentaries	United States	s1	Movie	September 25, 2021	2020	PG-13	80-100	2021-09-25	
1	Blood & Water	Ama Qamata	Unknown Director	International TV Shows	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	2021-09-24	
2	Blood & Water	Ama Qamata	Unknown Director	TV Dramas	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	2021-09-24	
3	Blood & Water	Ama Qamata	Unknown Director	TV Mysteries	South Africa	s2	TV Show	September 24, 2021	2021	TV-MA	2 Seasons	2021-09-24	
4	Blood &	Khosi	Unknown	International	South		TV	September			2		

```
#number of distinct titles on the basis of country
df_final1.groupby(['country']).agg({"title":"nunique"})
```

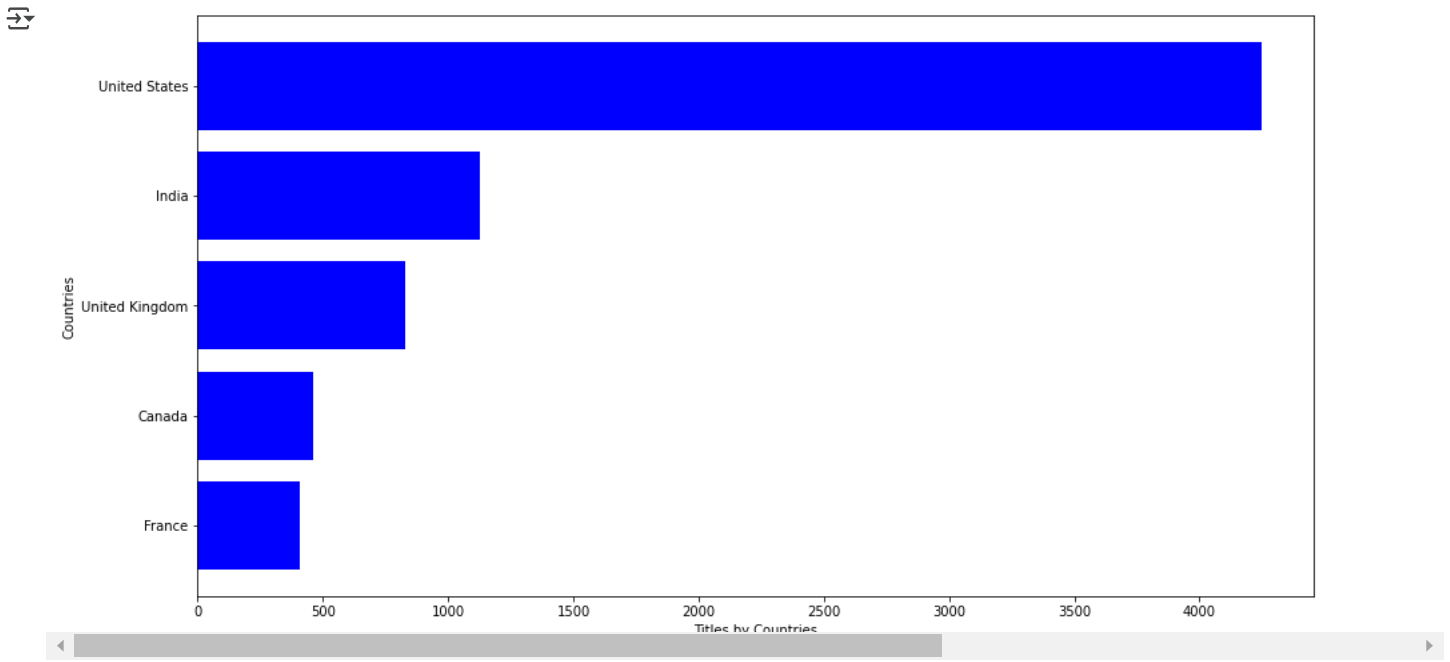
country	title
	3
Afghanistan	1
Albania	1
Algeria	3
Angola	2
Argentina	94
Armenia	1
Australia	162
Austria	12
Azerbaijan	1
Bahamas	1
Bangladesh	4
Belarus	1
Belgium	94
Bermuda	1
Botswana	1
Brazil	103
Bulgaria	10
Burkina Faso	1
Cambodia	6
Cameroon	2
Canada	460
Cayman Islands	2
Chile	30
China	166
Colombia	54
Croatia	4
Cuba	2
Cyprus	1
Czech Republic	23
Denmark	50
Dominican Republic	1
East Germany	1
Ecuador	1
Egypt	134
Ethiopia	1
Finland	12
France	409
Georgia	2
Germany	231
Ghana	8
Greece	11
Guatemala	2
Hong Kong	110
Hungary	11

Iceland	11
India	1126
Indonesia	97
Iran	4
Iraq	2
Ireland	46
Israel	30
Italy	102
Jamaica	1
Japan	338
Jordan	10
Kazakhstan	1
Kenya	6
Kuwait	9
Latvia	1
Lebanon	33
Liechtenstein	1
Lithuania	1
Luxembourg	12
Malawi	1
Malaysia	26
Malta	3
Mauritius	3
Mexico	175
Mongolia	1
Montenegro	1
Morocco	6
Mozambique	1
Namibia	2
Nepal	2
Netherlands	50
New Zealand	33
Nicaragua	1
Nigeria	140
Norway	30
Pakistan	24
Palestine	1
Panama	1
Paraguay	1
Peru	11
Philippines	90
Poland	42
Portugal	6
Puerto Rico	1
Qatar	10
Romania	14
Russia	27

Russia	27
Samoa	1
Saudi Arabia	14
Senegal	3
Serbia	7
Singapore	41
Slovakia	1
Slovenia	3
Somalia	1
South Africa	65
South Korea	235
Soviet Union	3
Spain	239
Sri Lanka	1
Sudan	1
Sweden	44
Switzerland	19
Syria	3
Taiwan	94
Thailand	74
Turkey	115
Uganda	1
Ukraine	3
United Arab Emirates	38
United Kingdom	831
United States	4246
Unknown Country	175
Uruguay	14
Vatican City	1
Venezuela	4
Vietnam	7
West Germany	5
Zimbabwe	3

Now it looks great.

```
df_country=df_final1.groupby(['country']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:5]
plt.figure(figsize=(15,8))
plt.barh(df_country[::-1]['country'], df_country[::-1]['title'],color=['blue'])
plt.xlabel('Titles by Countries')
plt.ylabel('Countries')
plt.show()
```

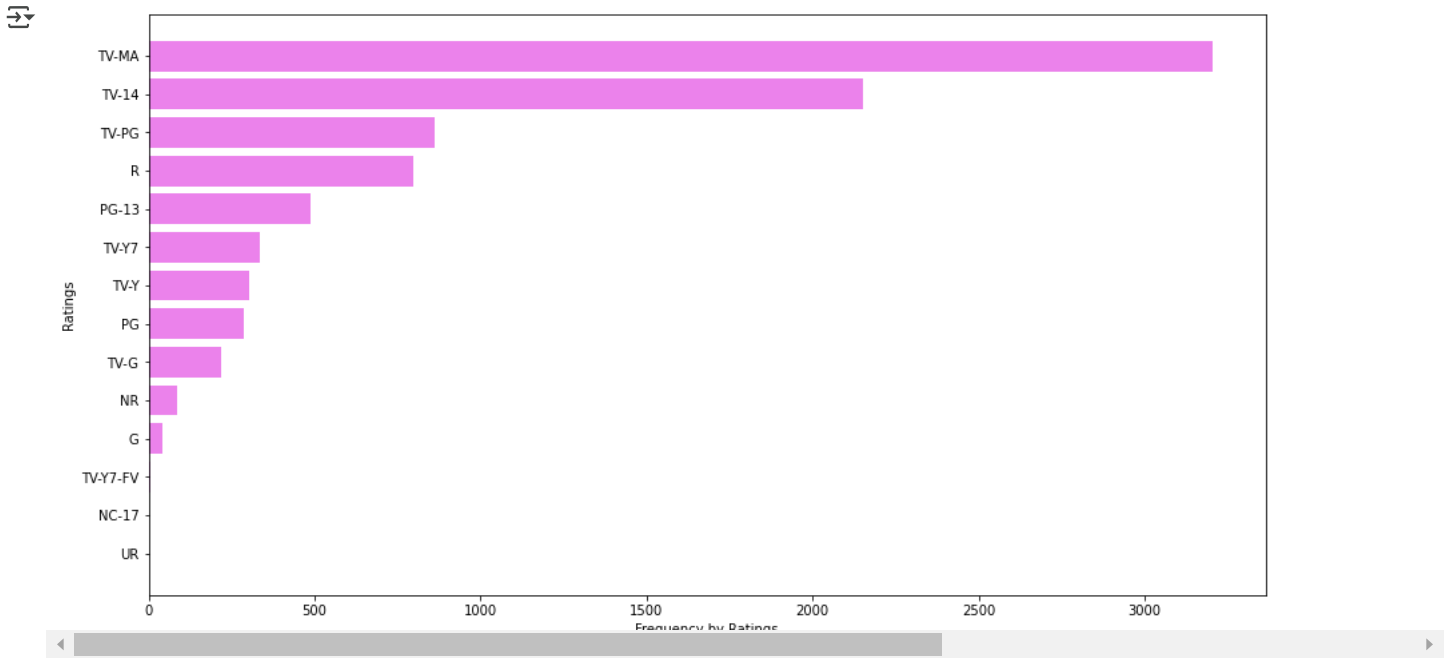


US,India,UK,Canada and France are leading countries in Content Creation on Netflix

```
#number of distinct titles on the basis of rating
df_final1.groupby(['rating']).agg({"title":"nunique"})
```

rating	title
G	41
NC-17	3
NR	87
PG	287
PG-13	490
R	799
TV-14	2151
TV-G	220
TV-MA	3204
TV-PG	863
TV-Y	305
TV-Y7	334
TV-Y7-FV	6
UR	3

```
df_rating=df_final1.groupby(['rating']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_rating[::-1]['rating'], df_rating[::-1]['title'],color=['violet'])
plt.xlabel('Frequency by Ratings')
plt.ylabel('Ratings')
plt.show()
```

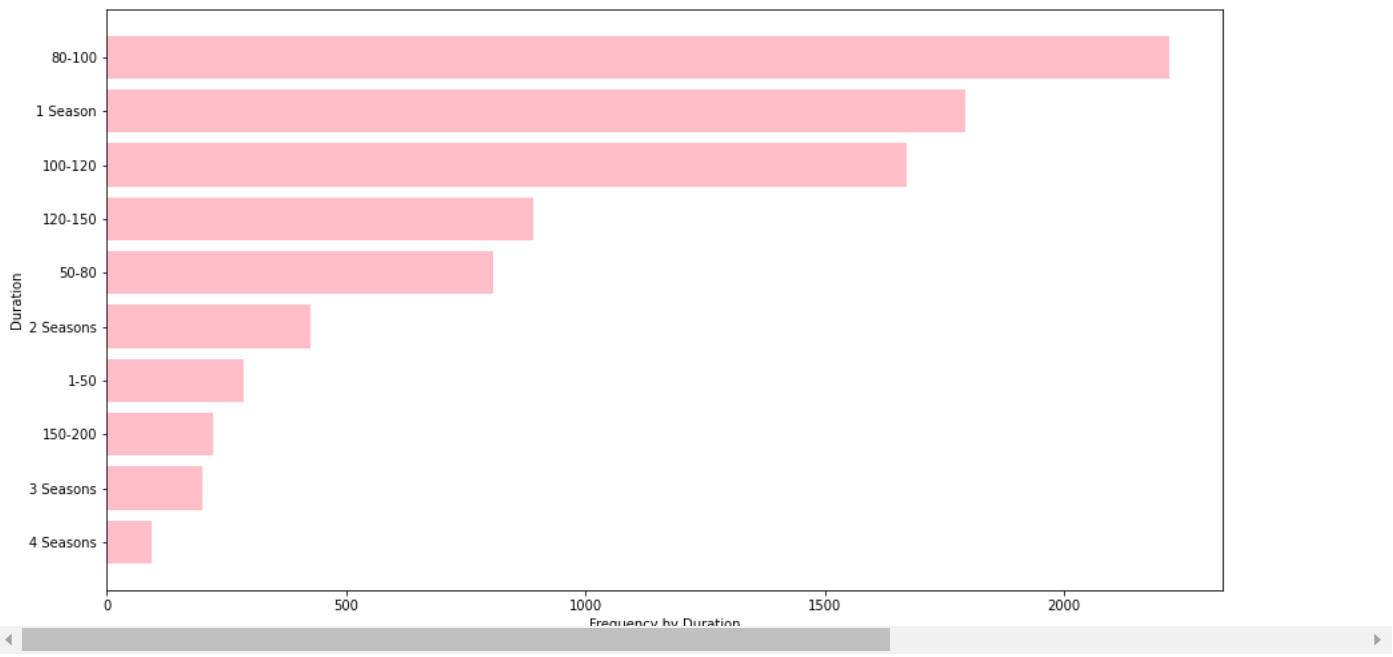


Most of the highly rated content on Netflix is intended for Mature Audiences, R Rated, content not intended for audience under 14 and those which require Parental Guidance

```
#number of distinct titles on the basis of duration
df_final1.groupby(['duration']).agg({"title":"nunique"})
```

duration	title
1 Season	1793
1-50	287
10 Seasons	7
100-120	1671
11 Seasons	2
12 Seasons	2
120-150	891
13 Seasons	3
15 Seasons	2
150-200	222
17 Seasons	1
2 Seasons	425
200-315	19
3 Seasons	199
4 Seasons	95
5 Seasons	65
50-80	808
6 Seasons	33
7 Seasons	23
8 Seasons	17
80-100	2220
9 Seasons	9

```
df_duration=df_final1.groupby(['duration']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
plt.figure(figsize=(15,8))
plt.barh(df_duration[::-1]['duration'], df_duration[::-1]['title'],color=['pink'])
plt.xlabel('Frequency by Duration')
plt.ylabel('Duration')
plt.show()
```



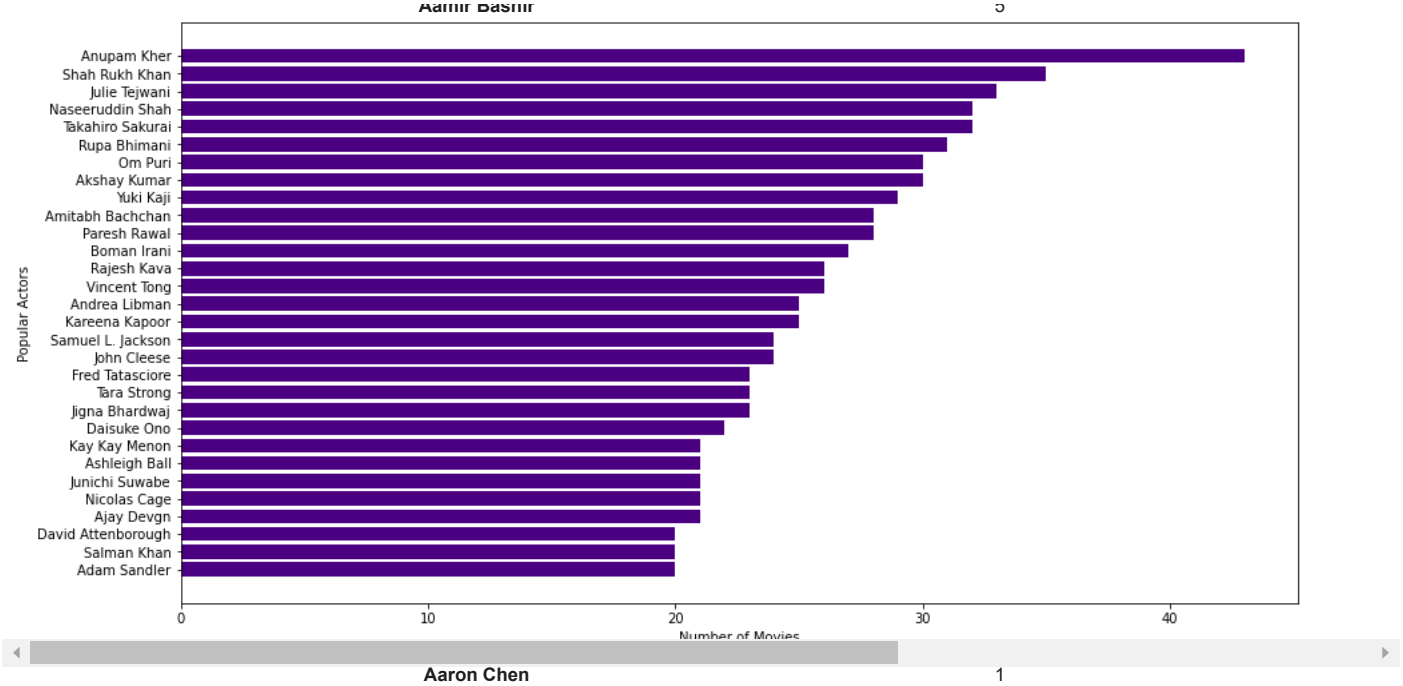
The duration of Most Watched content in our whole data is 80-100 mins.These must be movies and Shows having only 1 Season.

```
#number of distinct titles on the basis of Actors
df_final1.groupby(['Actors']).agg({"title":"nunique"})
```

Actors	title
Jr.	2
"Riley" Lakdhar Dridi	1
'Najite Dede	2
2 Chainz	1
2Mex	1
4Minute	1
50 Cent	5
9m88	1
A Boogie Wit tha Hoodie	1
A. Murat Özgen	1
A.C. Peterson	1
A.D. Miles	3
A.J. Cook	2
A.J. Johnson	1
A.J. LoCascio	3
A.K. Hangal	4
A.R. Rahman	1
A.S. Sasi Kumar	1
AC Lim	1
AFRA	1
AJ Bowen	1
AJ Michalka	1
AJ Rivera	1
ARAH	2
Aabhas Yadav	1
Aachal Munjal	1
Aadarsh Balakrishna	2
Aadhi	1
Aadhya Anand	1
Aadil Khan	1

```
df_actors=df_final1.groupby(['Actors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:31]
df_actors=df_actors[df_actors['Actors']!='Unknown Actor']
plt.figure(figsize=(15,8))
plt.barh(df_actors[::-1]['Actors'], df_actors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Actors')
plt.show()
```

Aakarshan Singh	1
Aakash Dabhade	4
Aakash Dahiya	1
Aakash Pandey	1
Aakshath Das	1
Aamina Sheikh	1
Aamir Ahmed	1
Aamir Bashir	5



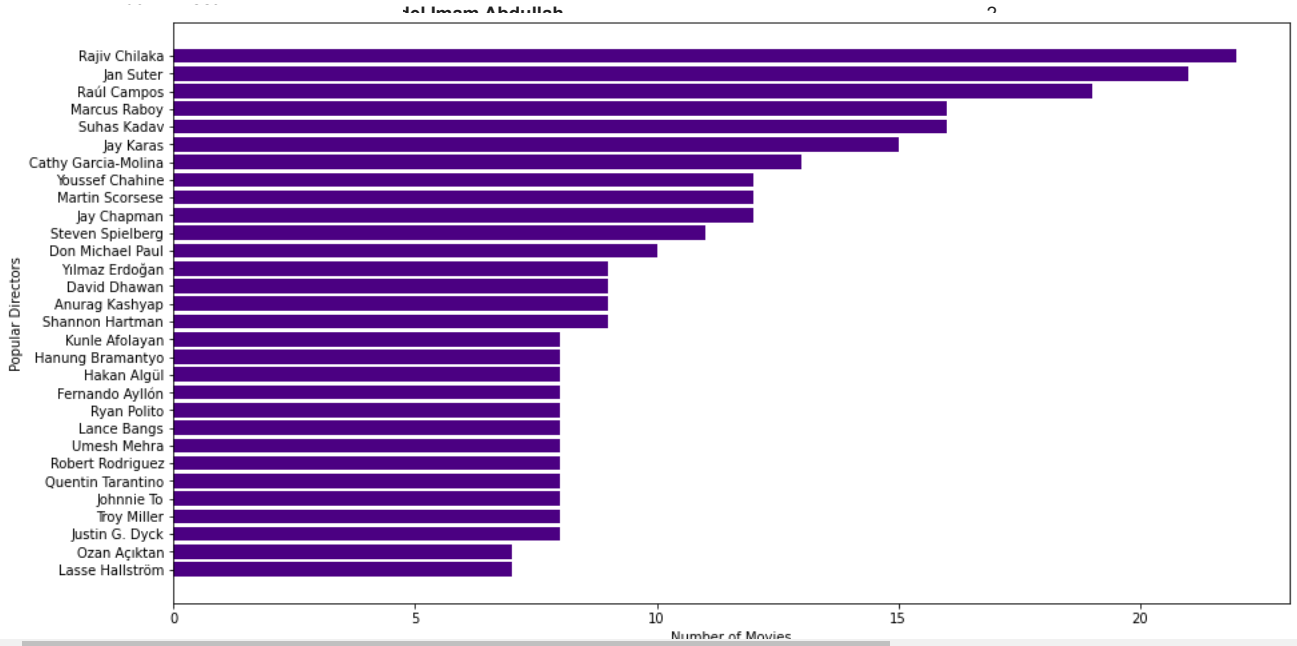
Anupam Kher,SRK,Julie Tejwani, Naseeruddin Shah and Takahiro Sakurai occupy the top stop in Most Watched content.

```
#number of distinct titles on the basis of Actors
df_final1.groupby(['Directors']).agg({"title":"nunique"})
```

Aaron Douglas	3
Aaron Eckhart	7
Aaron Eisenberg	1
Aaron Farb	1
Aaron Glenane	1
Aaron Guy	1
Aaron Hale	1
Aaron Hernandez	1
Aaron Hilmer	1
Aaron Himelstein	2
Aaron Jakubenko	3
Aaron Jeffery	3
Aaron Keogh	1
Aaron Kwok	1
Aaron L. McGrath	2
Aaron Marsden	2
Aaron Marshall	1
Aaron McCusker	2
Aaron Merke	1
Aaron Michael Drozin	1
Aaron Moorhead	1
Aaron Moten	1
Aaron Munoz	1
Aaron Paul	8
Aaron Pearl	1
Aaron Pedersen	1
Aaron Pruner	1

Directors	title
A. L. Vijay	2
A. Raajdheep	1
A. Salaam	1
A.R. Murugadoss	2
Aadish Keluskar	1
Aamir Bashir	1
Aamir Khan	1
Aanand Rai	1
Aaron Burns	1
Aaron Hancox	1
Aaron Hann	1
Aaron Lieber	1
Aaron Moorhead	2
Aaron Nee	1
Aaron Sorkin	2
Aaron Woodley	1
Aaron Woolf	1
Aatmaram Dharne	1
Abba T. Makama	2
Abbas Alibhai Burmawalla	5
Abbas Mustan	1
Abbas Tyrewala	1
Abby Epstein	1
Abdellatif Kechiche	1
Abdul Aziz Hashad	1
Abdulaziz Alshlahei	1
Abdullah Al Noor	1
Abel Ferrara	1
Abhay Chopra	1
Abhijeet Deshpande	1
Abhishek Chaubey	3
Abhishek Kapoor	3
Abhishek Saxena	1
Abhishek Sharma	3
Abhishek Varman	1
Abir Sengupta	1
Abosi Ogba	1
Abu Bakr Shawky	1
Achille Brice	1
Aco Tenriyagelli	1
Adam Alleca	1
Aaron Stanford	2
Aaron Staton	1
Aaron Taylor-Johnson	4
Aaron Tveit	1
Aaron Washington	1
Aaron Wolff	1
Aaron Yan	4
Aaron Yoo	3
Aarti Chhabria	1
Aarti Mann	1
Aarti Patel	1
Aarubala	1
Aarushi Sharma	1
Aarya Dave	1
Aarya DharmChand Kumar	1
Aaryan Menon	1
Aaryansh Malviya	1
Aarón Díaz	1
Aasha Pawar	1
Aashay Kulkarni	1
Aashi Rawal	1
Aashif Sheikh	1
Aashish Chaudhary	3
Aashish Kulkarni	1
Aasif Mandvi	6
Aayam Mehta	2
Aayan Boradia	1
Abayomi Alvin	1
Abba Ali Zaky	1
Abbas	1
Abbey Lee	1
Abbi Jacobson	3
Abby Quinn	1
Abby Rakic-Platt	1
Abby Ryder Fortson	1
Abby Trott	6
Abdalah Mishrif	4
Abdalla Mahmoud	2
Abdel Aziz El Mountassir	1
Abdel Aziz Khalil	1
Abdel Ghani Benizza	1
Abdel Ghani Nagdi	1

```
df_directors=df_final1.groupby(['Directors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:31]
df_directors=df_directors[df_directors['Directors']!='Unknown Director']
plt.figure(figsize=(15,8))
plt.barh(df_directors[::-1]['Directors'], df_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Directors')
plt.show()
```

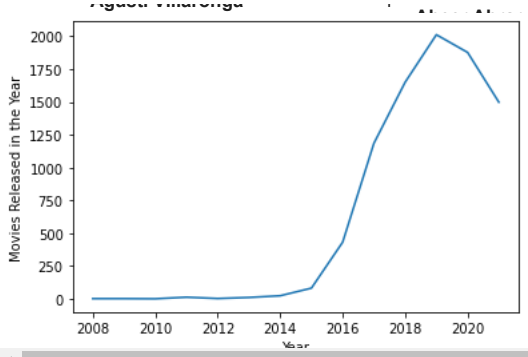


Rajiv Chilaka, Jan Suter and Raul Campos are the most popular directors across Netflix

```
#number of distinct titles on the basis of year
df_final1.groupby(['year']).agg({"title":"nunique"})
```

year	title
2008	2
2009	2
2010	1
2011	13
2012	3
2013	11
2014	24
2015	82
2016	432
2017	1185
2018	1650
2019	2012
2020	1877
2021	1498

```
df_year=df_final1.groupby(['year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_year, x='year', y='title')
plt.ylabel("Movies Released in the Year")
plt.xlabel("Year")
plt.show()
```

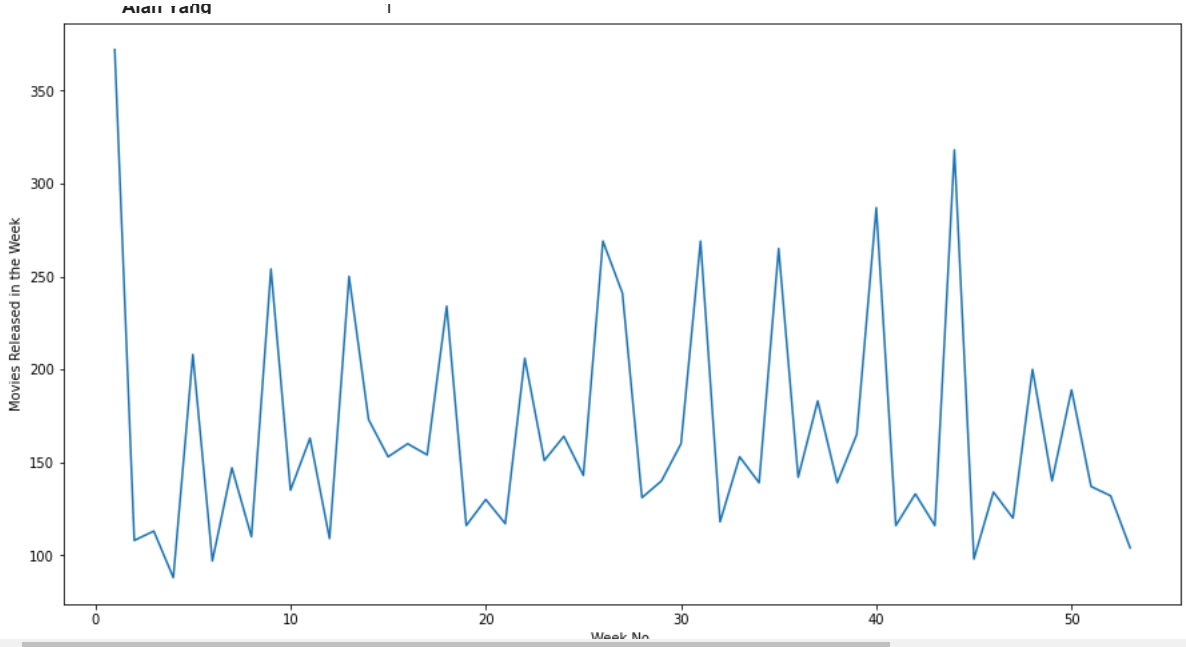


The Amount of Content across Netflix has increased from 2008 continuously till 2019. Then started decreasing from here(probably due to Covid)

```
#number of distinct titles on the basis of week
df_final1.groupby(['week_Added']).agg({"title":"nunique"})
```

week_Added	title
1	372
2	108
3	113
4	88
5	208
6	97
7	147
8	110
9	254
10	135
11	163
12	109
13	250
14	173
15	153
16	160
17	154
18	234
19	116
20	130
21	117
22	206
23	151
24	164
25	143
26	269
27	241
28	131
29	140
30	160

```
df_week=df_final1.groupby(['week_Added']).agg({"title":"nunique"}).reset_index()
plt.figure(figsize=(15,8))
sns.lineplot(data=df_week, x='week_Added', y='title')
plt.ylabel("Movies Released in the Week")
plt.xlabel("Week No.")
plt.show()
```

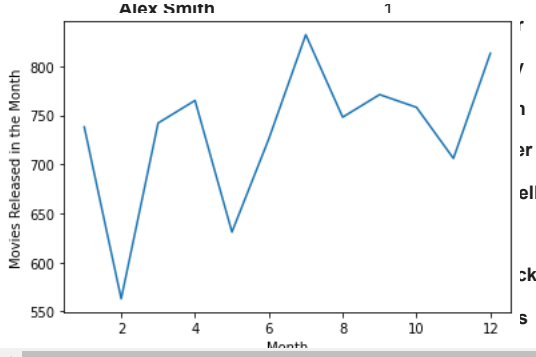


Most of the Content across Netflix is added in the first week of the year and it follows a bit of a cyclical pattern

```
#number of distinct titles on the basis of week
df_final1.groupby(['month_added']).agg({"title":"nunique"})
```

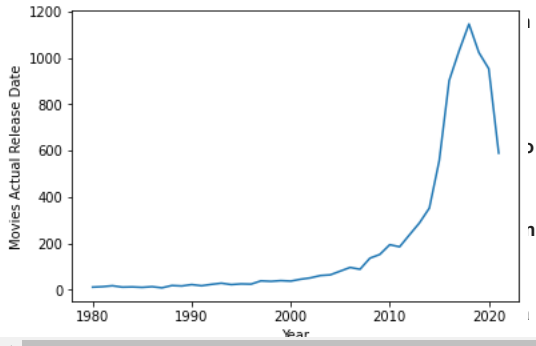
month_added	title
1	738
2	563
3	742
4	765
5	631
6	726
7	832
8	748
9	771
10	758
11	706
12	813

```
df_month=df_final1.groupby(['month_added']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_month, x='month_added', y='title')
plt.ylabel("Movies Released in the Month")
plt.xlabel("Month")
plt.show()
```



Most of the content is added in the first and last months across Netflix(reinstating what we observed for first week in baove plot)

```
df_release_year=df_final1[df_final1['release_year']>=1980].groupby(['release_year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_release_year, x='release_year', y='title')
plt.ylabel("Movies Actual Release Date")
plt.xlabel("Year")
plt.show()
```

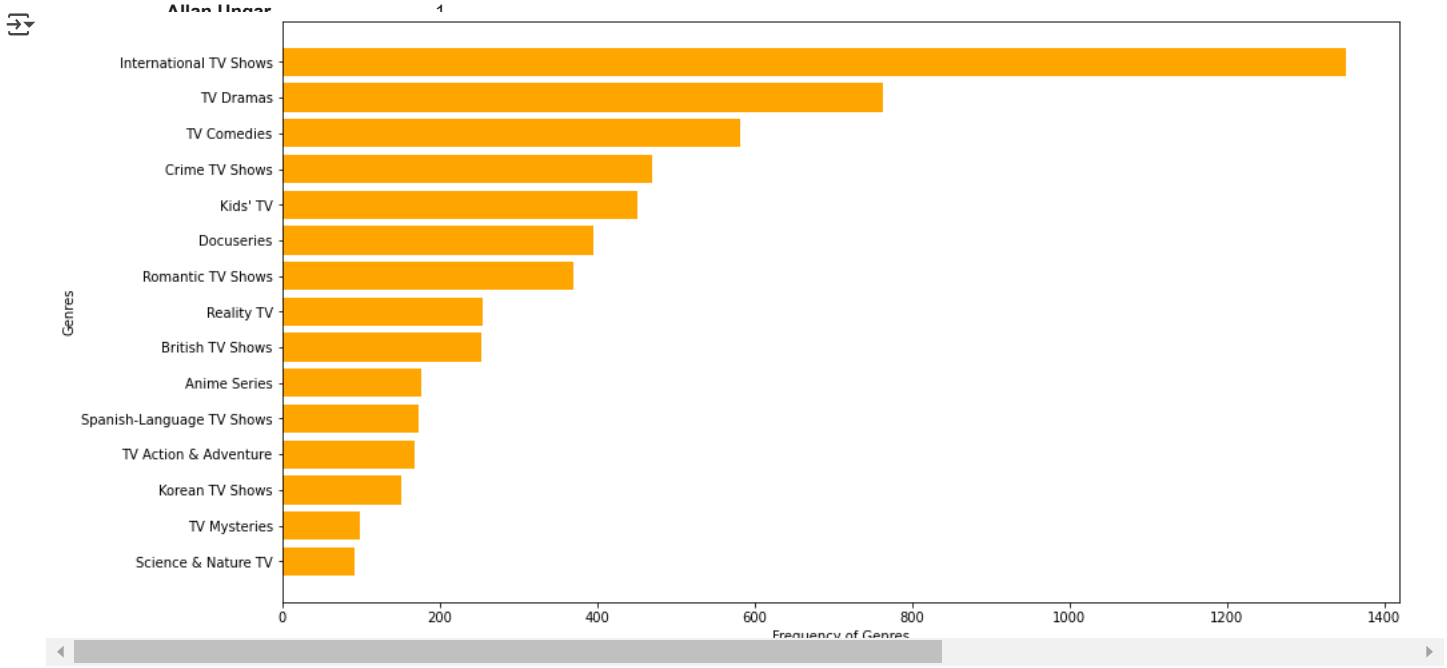


Net content release which are later uploaded to Netflix has increased since 1980 till 2020 though later reduced certainly due to COVID-19

Univariate Analysis separately for shows and movies

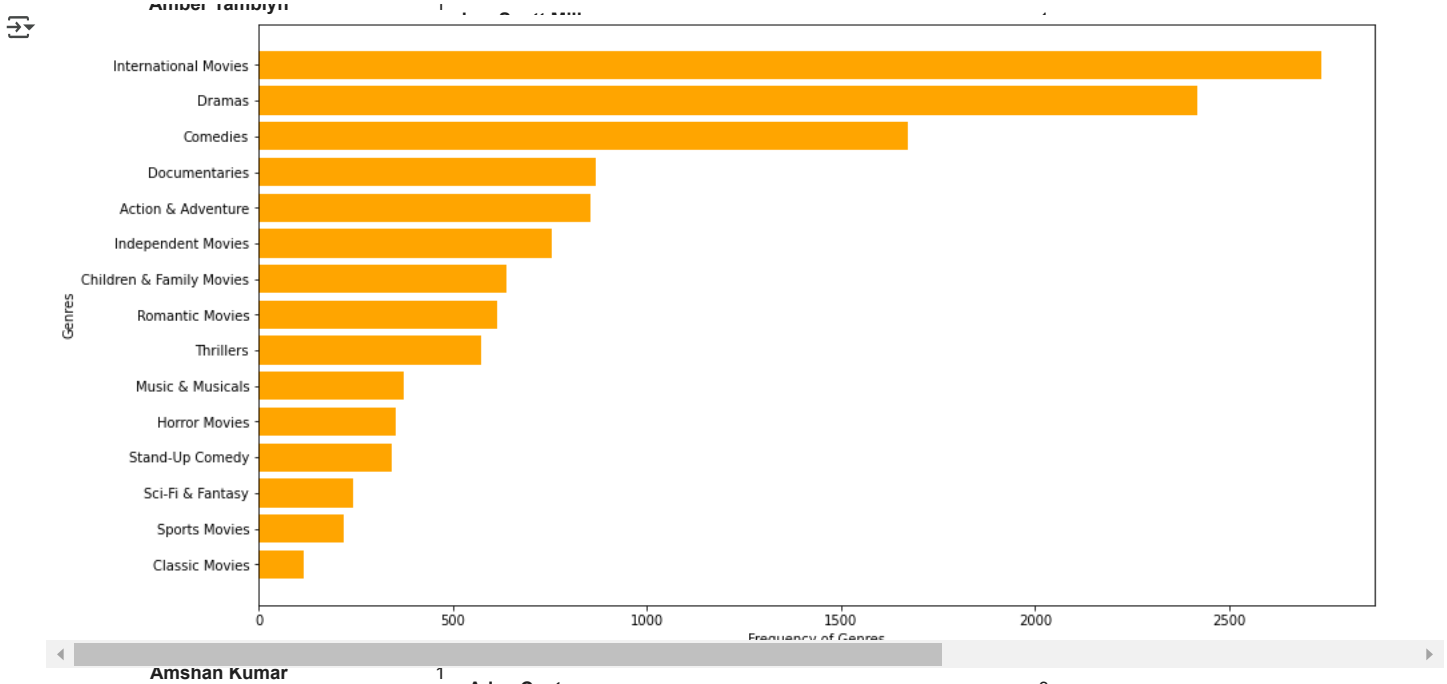
```
df_shows=df_final1[df_final1['type']=='TV Show']
df_movies=df_final1[df_final1['type']=='Movie']
```

```
df_genre=df_shows.groupby(['Genre']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_genre[::-1]['Genre'], df_genre[::-1]['title'],color=['orange'])
plt.xlabel('Frequency of Genres')
plt.ylabel('Genres')
plt.show()
```



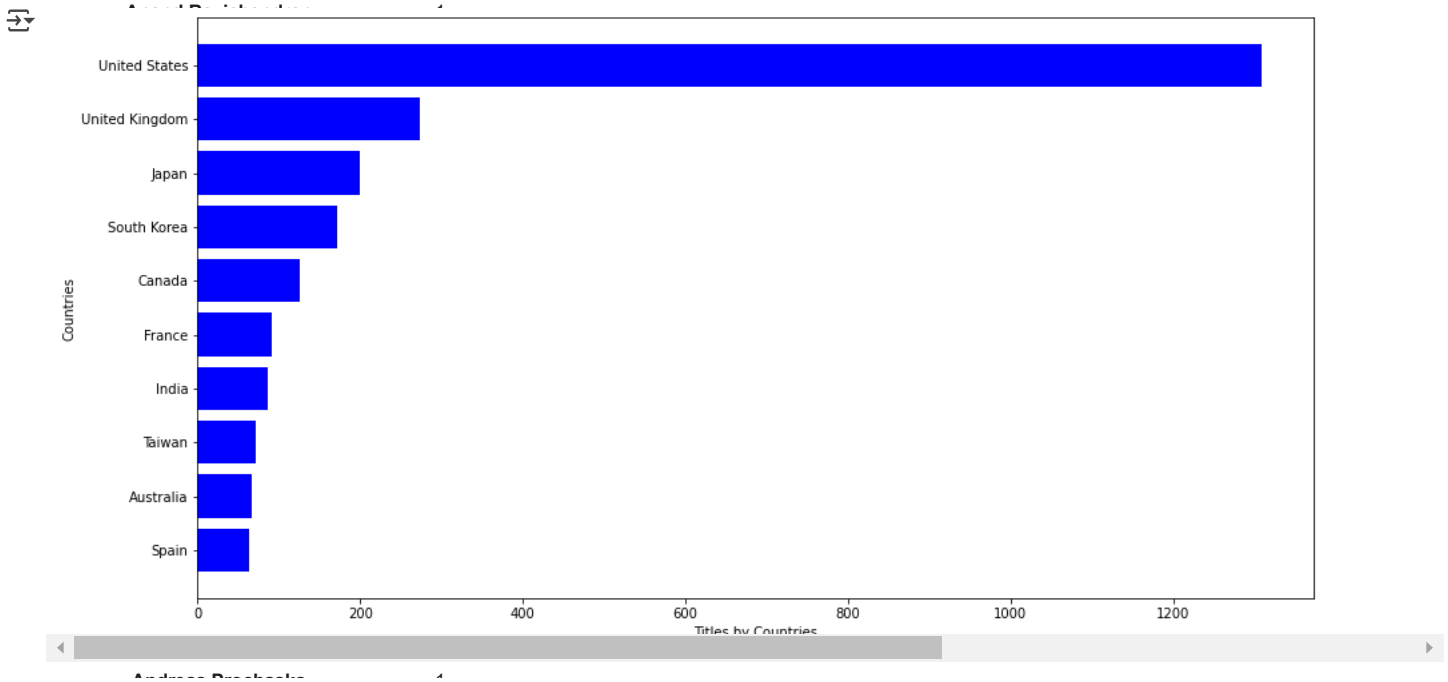
International TV Shows, Dramas and Comedy Genres are popular across TV Shows in Netflix

```
df_genre=df_movies.groupby(['Genre']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_genre[::-1]['Genre'], df_genre[::-1]['title'],color=['orange'])
plt.xlabel('Frequency of Genres')
plt.ylabel('Genres')
plt.show()
```

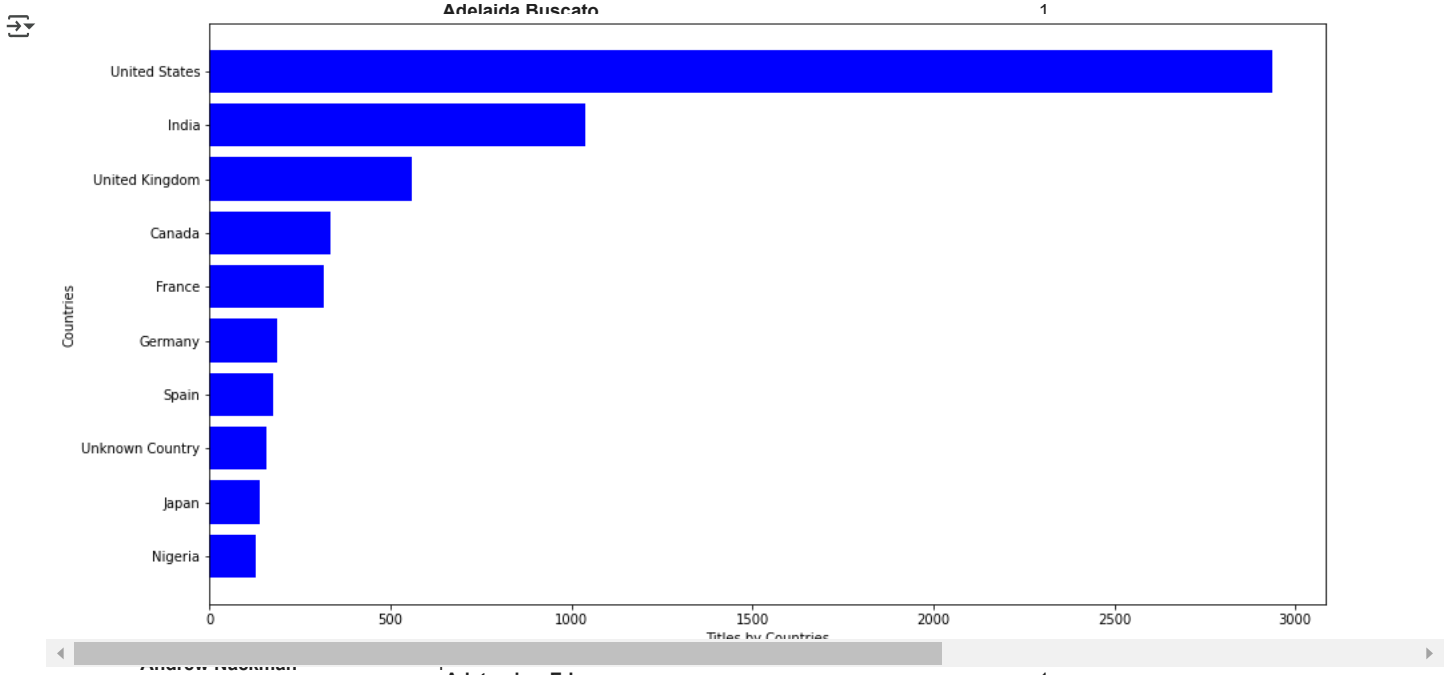


International Movies, Dramas and Comedy Genres are popular followed by Documentaries across Movies on Netflix

```
df_country=df_shows.groupby(['country']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
plt.figure(figsize=(15,8))
plt.barh(df_country[::-1]['country'], df_country[::-1]['title'],color=['blue'])
plt.xlabel('Titles by Countries')
plt.ylabel('Countries')
plt.show()
```



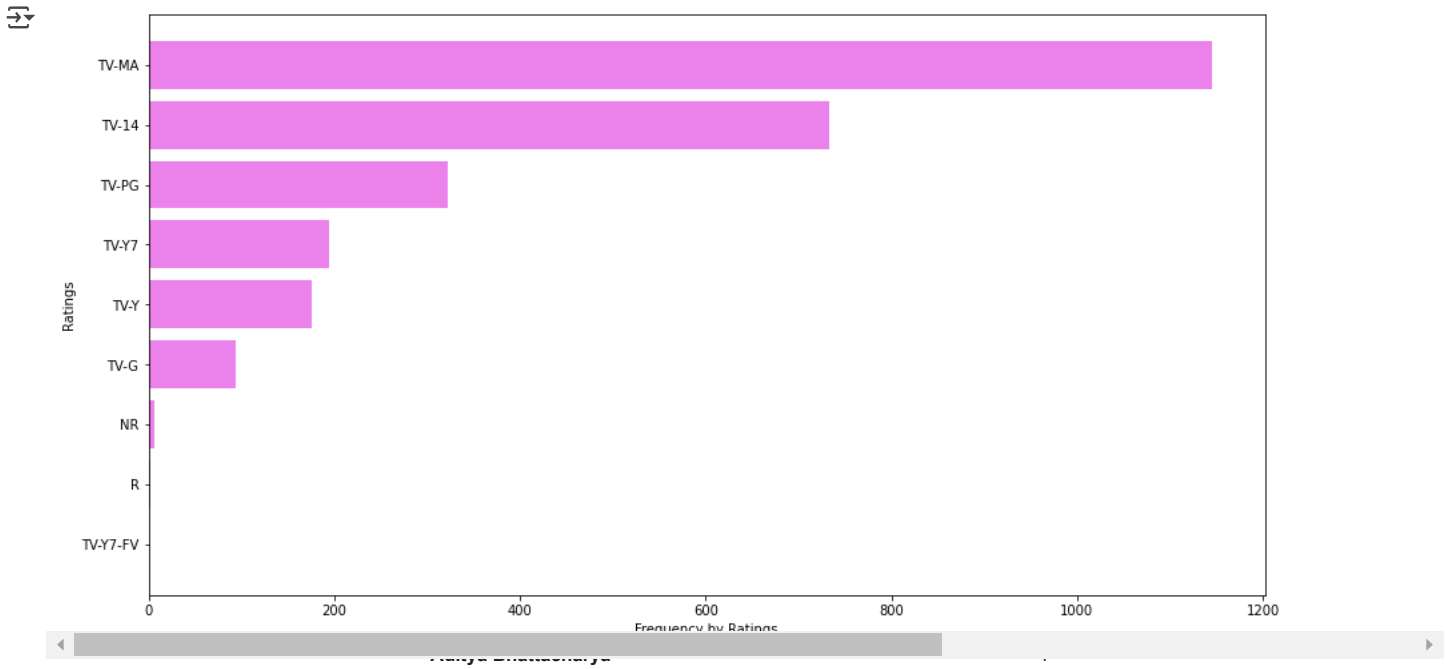

```
df_country=df_movies.groupby(['country']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
plt.figure(figsize=(15,8))
plt.barh(df_country[::-1]['country'], df_country[::-1]['title'],color=['blue'])
plt.xlabel('Titles by Countries')
plt.ylabel('Countries')
plt.show()
```



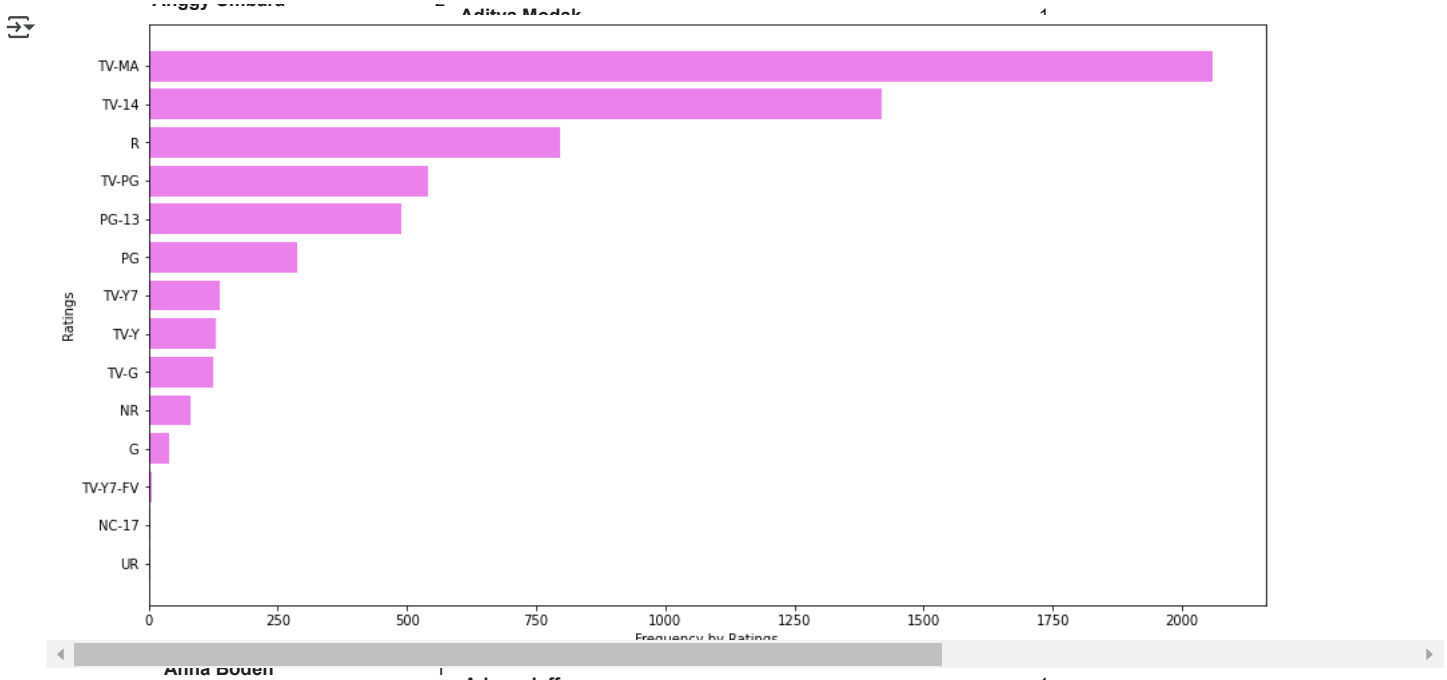
United States is leading across both TV Shows and Movies. UK also provides great content across TV Shows and Movies. Surprisingly India is much more prevalent in Movies as compared TV Shows.

Moreover the number of Movies created in India outweigh the sum of TV Shows and Movies across UK since India was rated as second in net sum of whole content across Netflix.

```
df_rating=df_shows.groupby(['rating']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_rating[::-1]['rating'], df_rating[::-1]['title'],color=['violet'])
plt.xlabel('Frequency by Ratings')
plt.ylabel('Ratings')
plt.show()
```



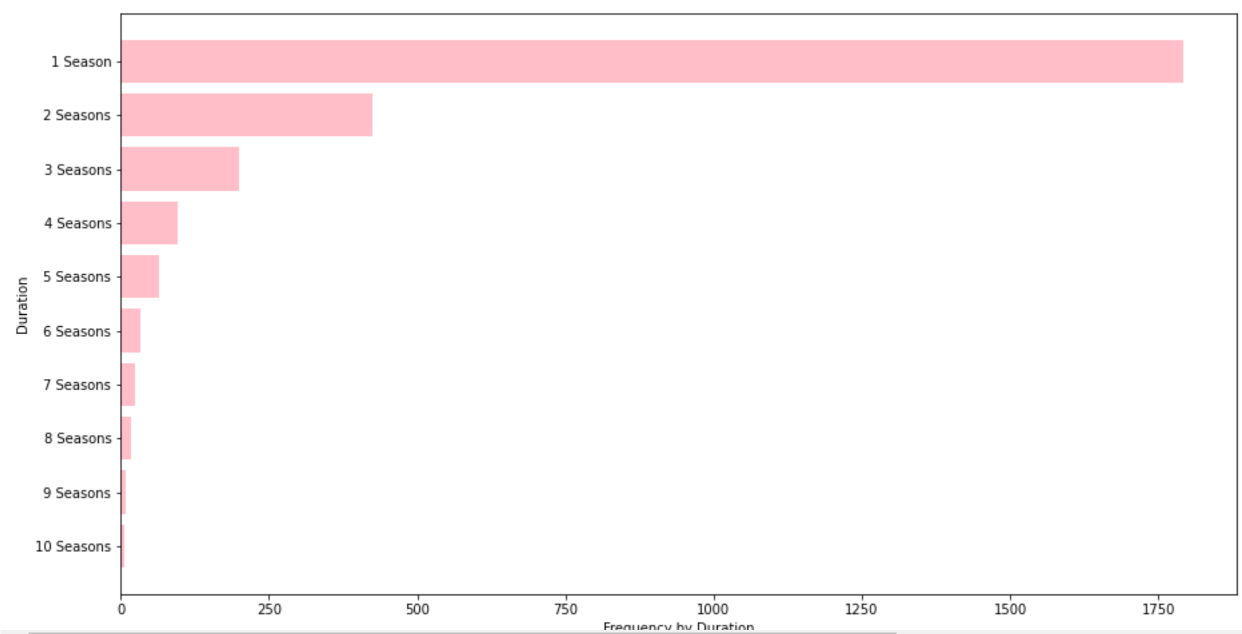

```
df_rating=df_movies.groupby(['rating']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_rating[::-1]['rating'], df_rating[::-1]['title'],color=['violet'])
plt.xlabel('Frequency by Ratings')
plt.ylabel('Ratings')
plt.show()
```



So it seems plausible to conclude that the popular ratings across Netflix includes Mature Audiences and those appropriate for over 14/over 17 ages.

Moreover there are no TV Shows having a rating of R

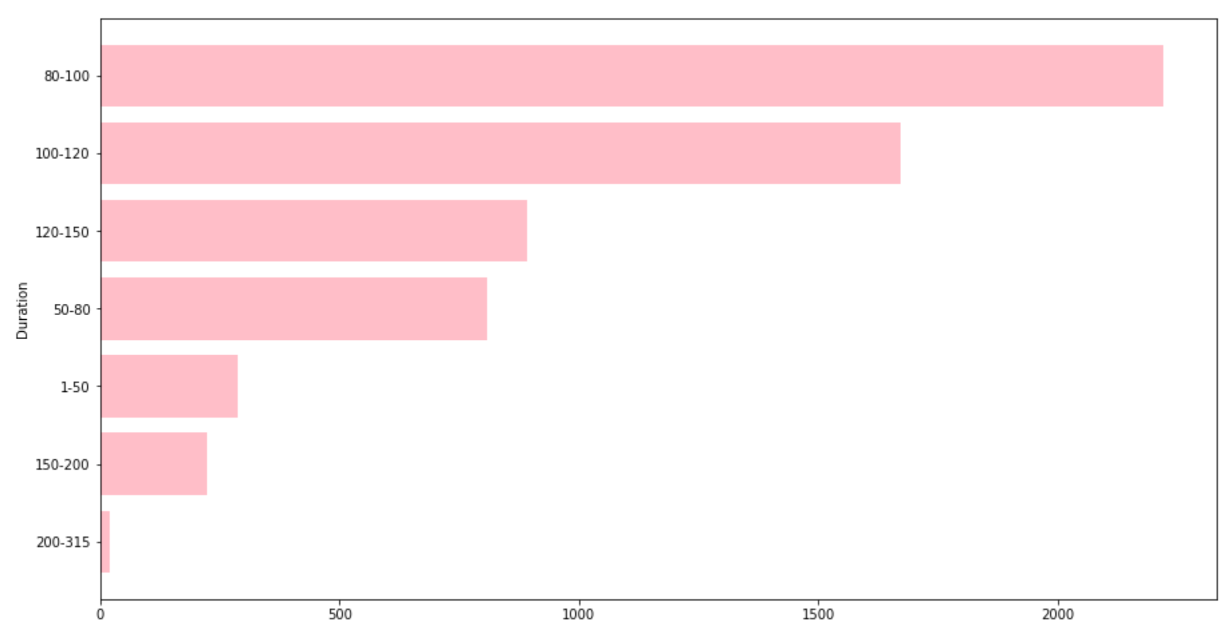
```
df_duration=df_shows.groupby(['duration']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
plt.figure(figsize=(15,8))
plt.barh(df_duration[::-1]['duration'], df_duration[::-1]['title'],color=['pink'])
plt.xlabel('Frequency by Duration')
plt.ylabel('Duration')
plt.show()
```



Across TV Shows, shows having only 1 Season are common as soon as the season length increases, the number of shows decrease and this definitely sounds as expected

Antoine Fuqua 4
Adrian Majewski 1
Antoinette Jadaone 5
Adrian Martinez 2
Anton Corbijn 1

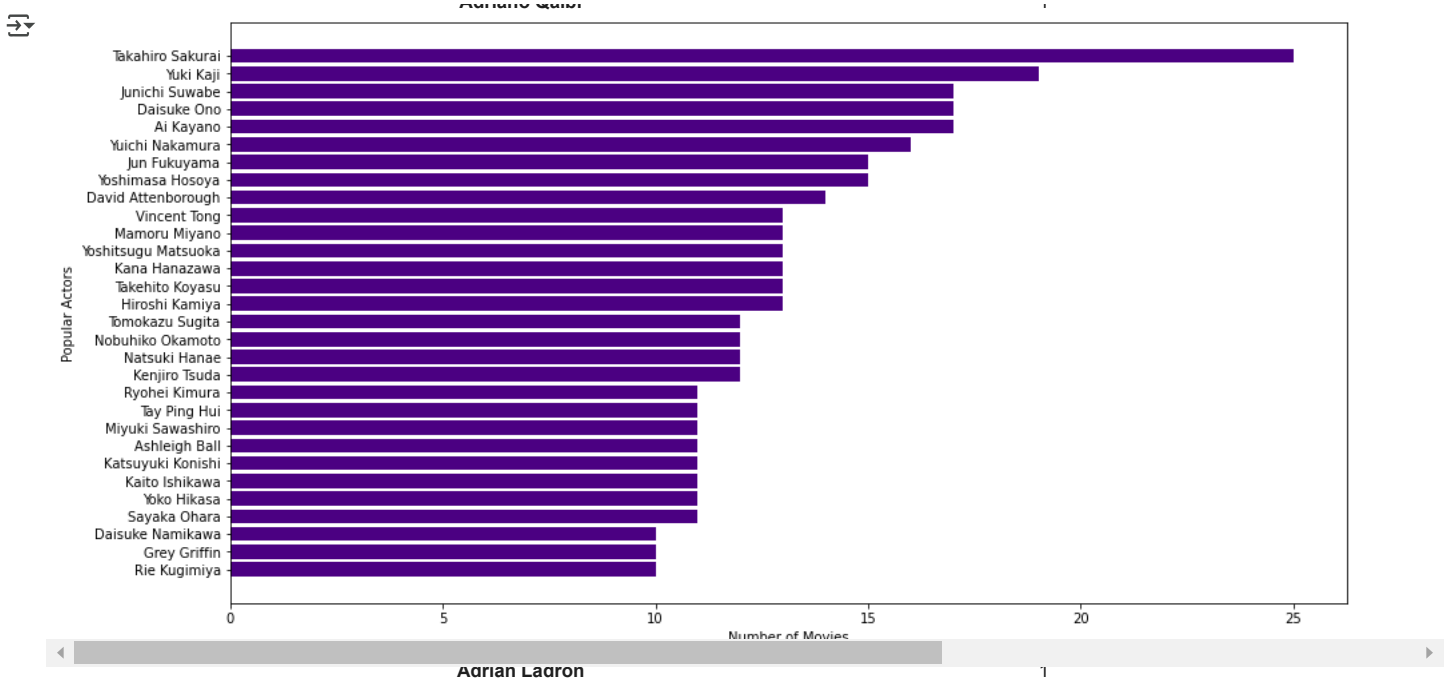
```
df_duration=df_movies.groupby(['duration']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
plt.figure(figsize=(15,8))
plt.barh(df_duration[::-1]['duration'], df_duration[::-1]['title'],color=['pink'])
plt.xlabel('Frequency by Duration')
plt.ylabel('Duration')
plt.show()
```



Across movies 80-100,100-120 and 120-150 is the ranges of minutes for which most movies lie. So quite possibly 80-150 mins is the sweet spot we would be wanting for movies.

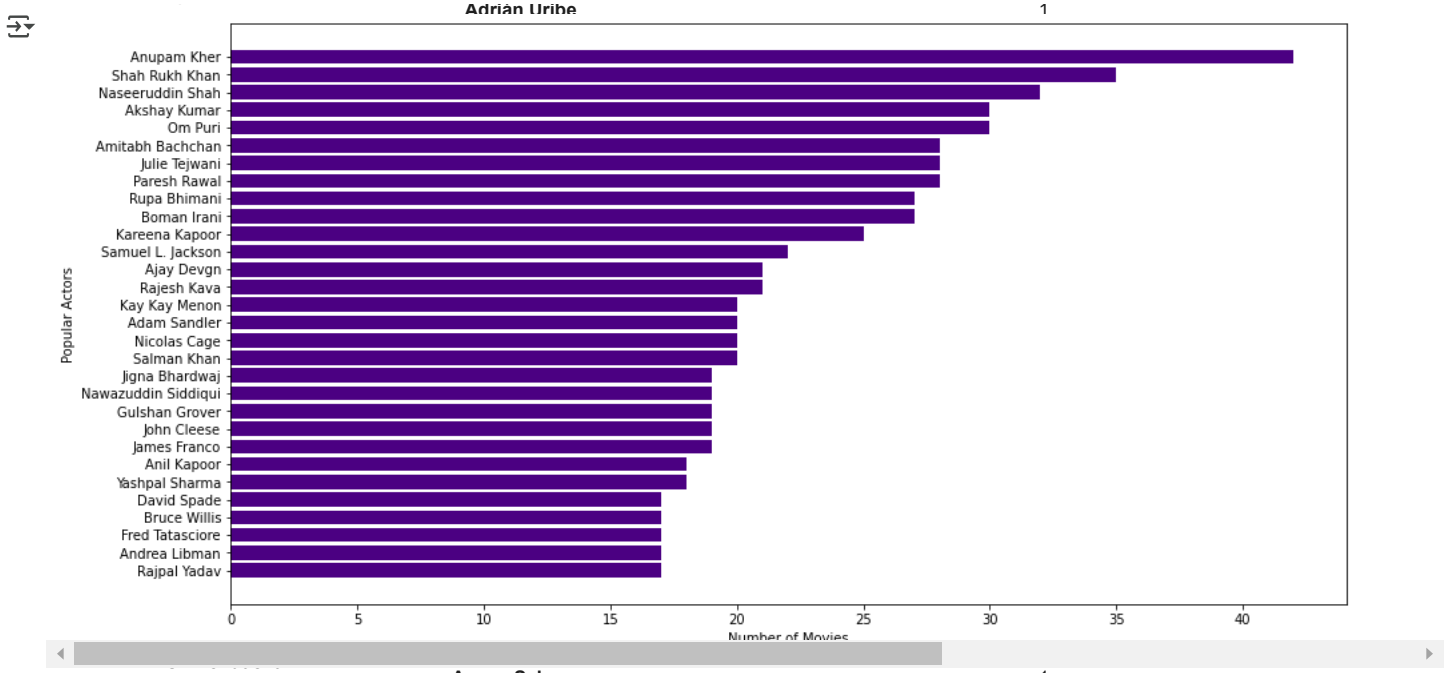
Appie Boudellah 1
April Mullen 1
Adrianna Chlebicka 1
Apurva Dhar Badgaiyann 1

```
df_actors=df_shows.groupby(['Actors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:31]
df_actors=df_actors[df_actors['Actors']!='Unknown Actor']
plt.figure(figsize=(15,8))
plt.barh(df_actors[::-1]['Actors'], df_actors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Actors')
plt.show()
```



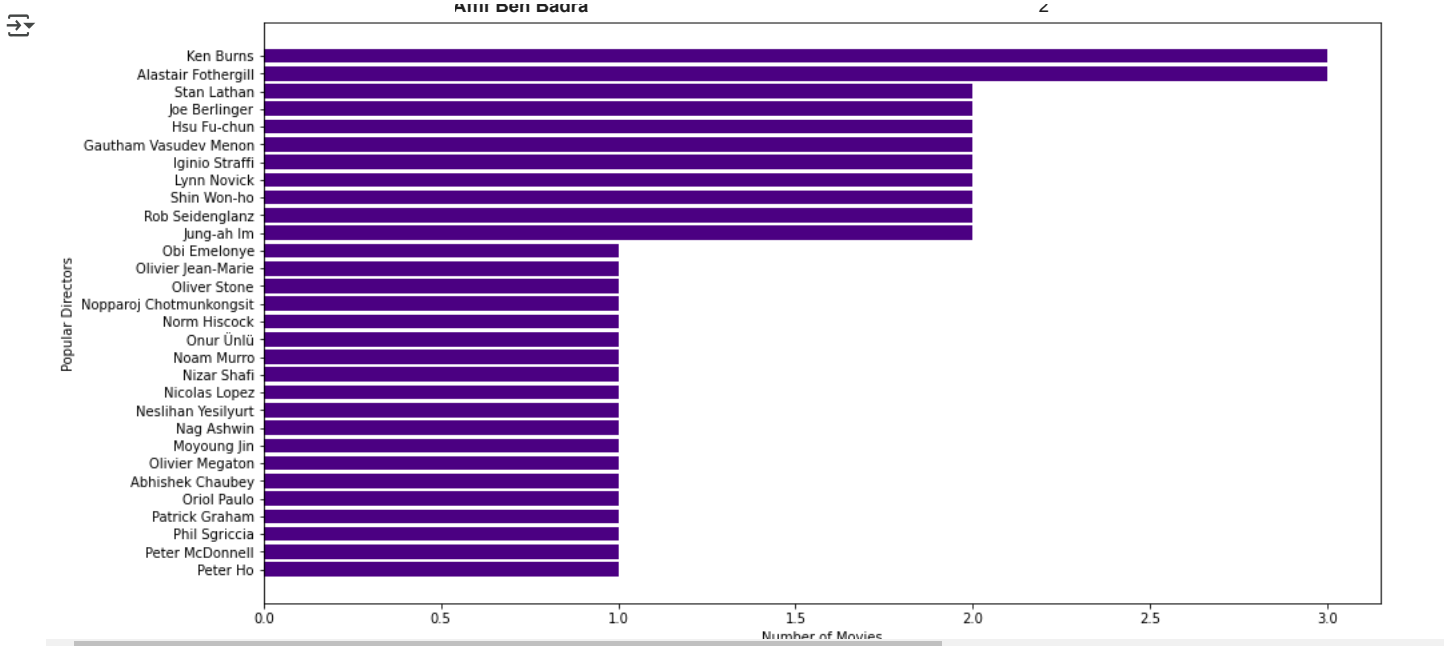
Takahiro Sakurai,Yuki Kaji and other South Korean/Japanese actors are the most popular actors across TV Shows

```
df_actors=df_movies.groupby(['Actors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:31]
df_actors=df_actors[df_actors['Actors']!='Unknown Actor']
plt.figure(figsize=(15,8))
plt.barh(df_actors[::-1]['Actors'], df_actors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Actors')
plt.show()
```



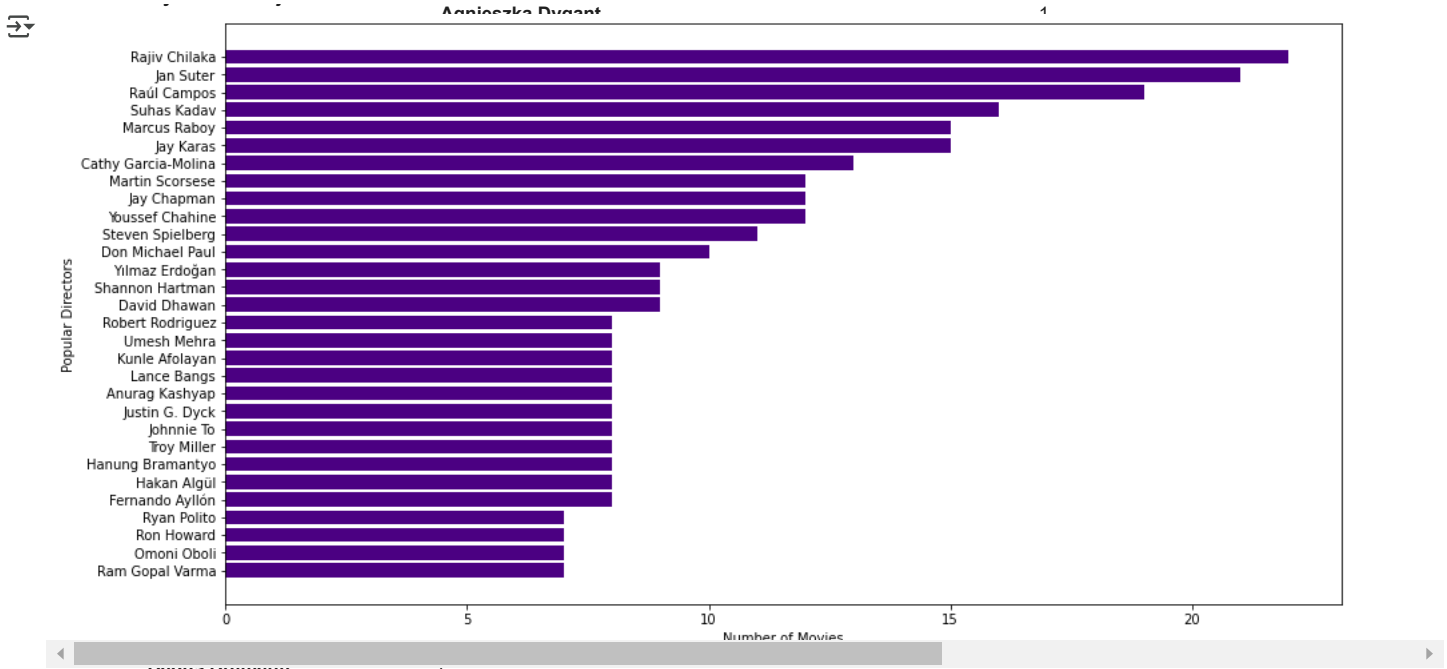
Our bollywood actors such as Anupam Kher, SRK, Naseeruddin Shah are very much popular acrossmovies on Netflix

```
df_directors=df_shows.groupby(['Directors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:31]
df_directors=df_directors[df_directors['Directors']!='Unknown Director']
plt.figure(figsize=(15,8))
plt.barh(df_directors[::-1]['Directors'], df_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Directors')
plt.show()
```



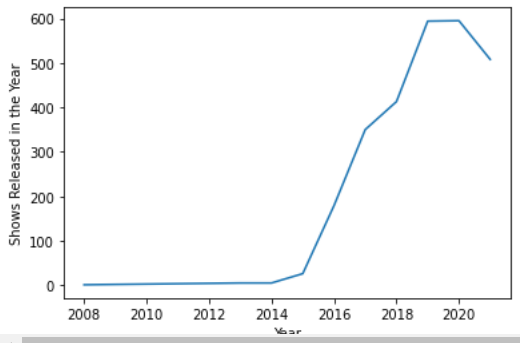
Ken Burns, Alastair Fothergill, Stan Lathan, Joe Barlinger are popular directors across TV Shows on Netflix

```
df_directors=df_movies.groupby(['Directors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:31]
df_directors=df_directors[df_directors['Directors']!='Unknown Director']
plt.figure(figsize=(15,8))
plt.barh(df_directors[::-1]['Directors'], df_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Directors')
plt.show()
```



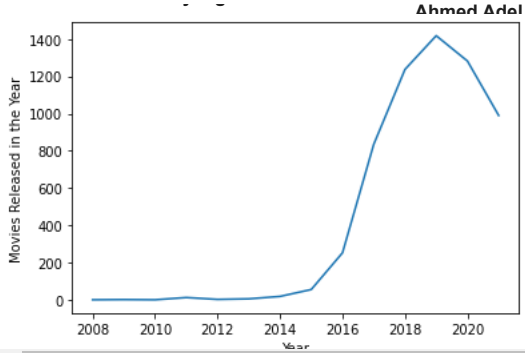
Rajiv Chilka, Jan Suter, Raúl Campos, Suhas Kadav are popular directors across movies

```
df_year=df_shows.groupby(['year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_year, x='year', y='title')
plt.ylabel("Shows Released in the Year")
plt.xlabel("Year")
plt.show()
```



Ahmad Thaher 1

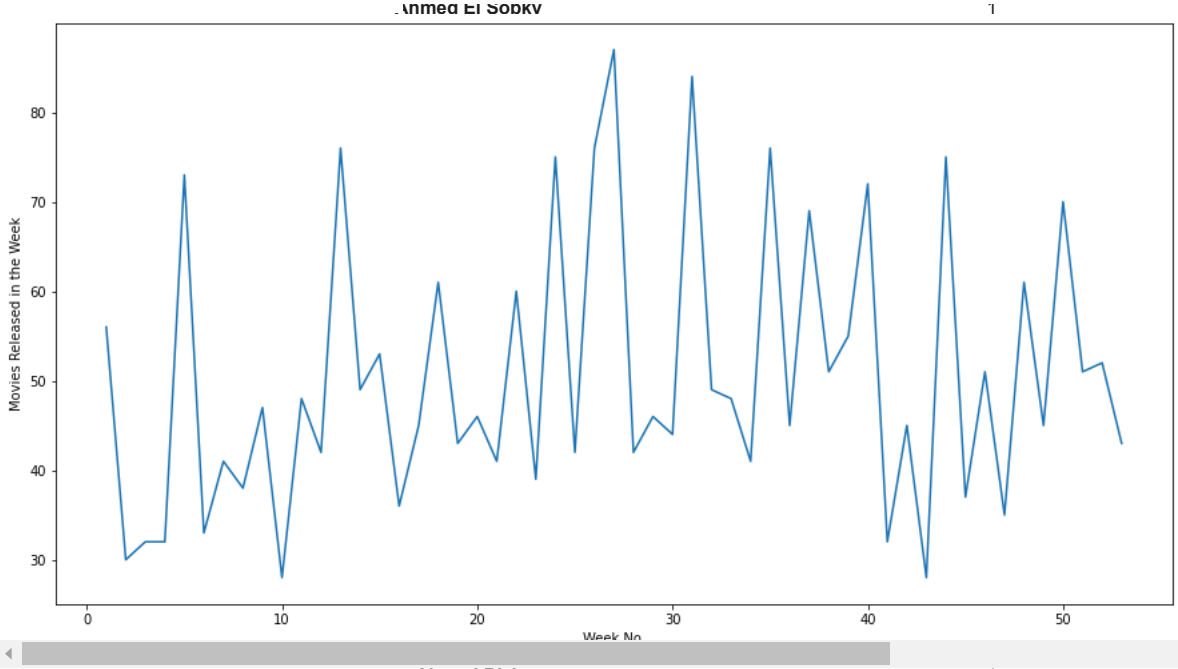
```
df_year=df_movies.groupby(['year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_year, x='year', y='title')
plt.ylabel("Movies Released in the Year")
plt.xlabel("Year")
plt.show()
```



Ahmed Dash 1
Ahmed Dawood 5
Ahmed Eid 3

Till 2019, overall content across Netflix was increasing but due to Covid in 2020, though TV Shows didn't take a hit then Movies did take a hit.
Well later in 2020 content across both was reduced significantly

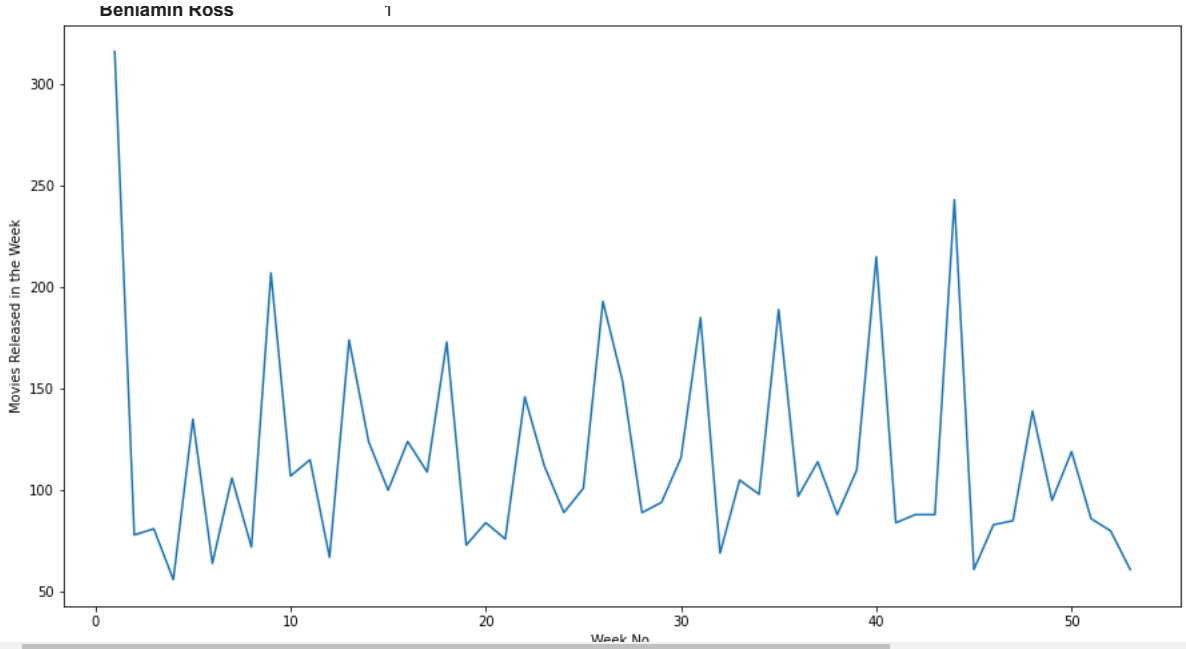
```
df_week=df_shows.groupby(['week_Added']).agg({"title":"nunique"}).reset_index()
plt.figure(figsize=(15,8))
sns.lineplot(data=df_week, x='week_Added', y='title')
plt.ylabel("Movies Released in the Week")
plt.xlabel("Week No.")
plt.show()
```



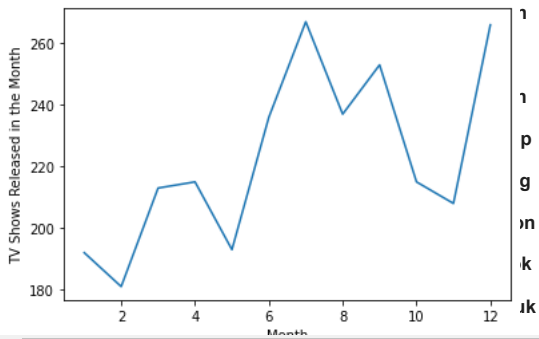
Ahmed Rizk 1

```
df_week=df_movies.groupby(['week_Added']).agg({"title":"nunique"}).reset_index()
plt.figure(figsize=(15,8))
```

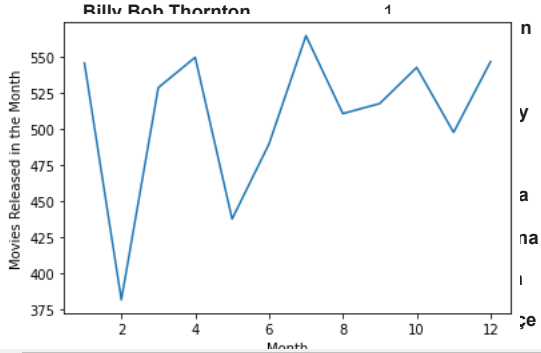
```
sns.lineplot(data=df_week, x='week_Added', y='title')
plt.ylabel("Movies Released in the Week")
plt.xlabel("Week No.")
plt.show()
```




```
df_month=df_shows.groupby(['month_added']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_month, x='month_added', y='title')
plt.ylabel("TV Shows Released in the Month")
plt.xlabel("Month")
plt.show()
```



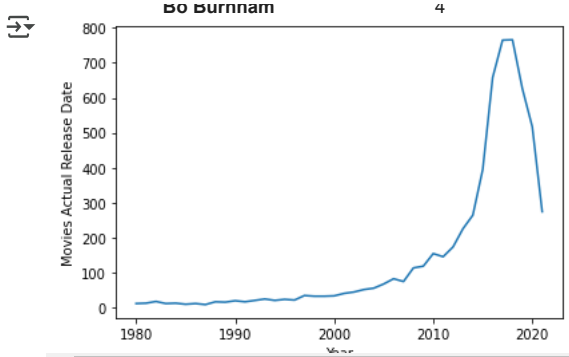

```
df_month=df_movies.groupby(['month_added']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_month, x='month_added', y='title')
plt.ylabel("Movies Released in the Month")
plt.xlabel("Month")
plt.show()
```



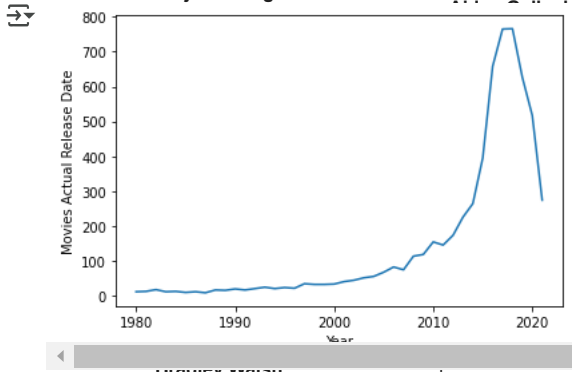
TV Shows are added in Netflix by a tremendous amount in mid weeks/months of the year, i.e- July

Movies are added in Netflix by a tremendous amount in first week/last month of current year and first month of next year

```
df_release_year=df_movies[df_movies['release_year']>=1980].groupby(['release_year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_release_year, x='release_year', y='title')
plt.ylabel("Movies Actual Release Date")
plt.xlabel("Year")
plt.show()
```




```
df_release_year=df_movies[df_movies['release_year']>=1980].groupby(['release_year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_release_year, x='release_year', y='title')
plt.ylabel("Movies Actual Release Date")
plt.xlabel("Year")
plt.show()
```



Actual Releases of both TV Shows and Movies have taken a hit after 2020

Questions to be Explored Now for Recommendations

1) So this time, the granularity level is country and analysis of TV Shows/Movies the country brings. I am going to consider only the top countries individually for TV Shows and Movies. There are definitely some common countries too which bring out quality content in both TV Shows and Movies.

2) Which Genres do these countries offer and what are the intended audiences(Ratings) which are popular in Netflix?

3)In case of Movies, what is the duration/length of movies which makes them special and depicts attention span?

4)Who are the popular actors/directors across TV Shows and Movies in these countries?

5)In what time of the year people tend to watch movies and shows in these countries?

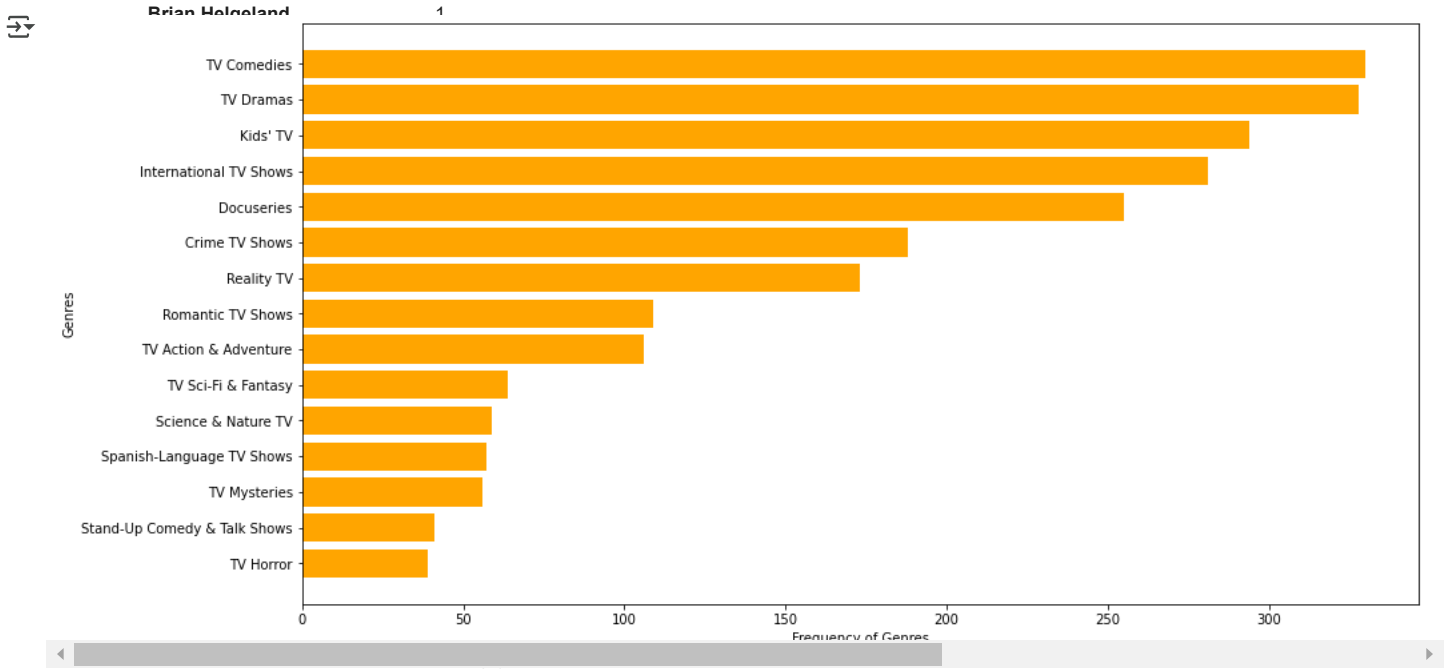
6)Popular Actor and Director combinations in these countries

```
#below countries will be analyzed for both shows and movies
shows_and_movies=['United States','India','United Kingdom']
#below countries will be only analyzed on basis of shows
only_shows=['Japan','South Korea']
```

Univariate Analysis separately for shows and movies in USA

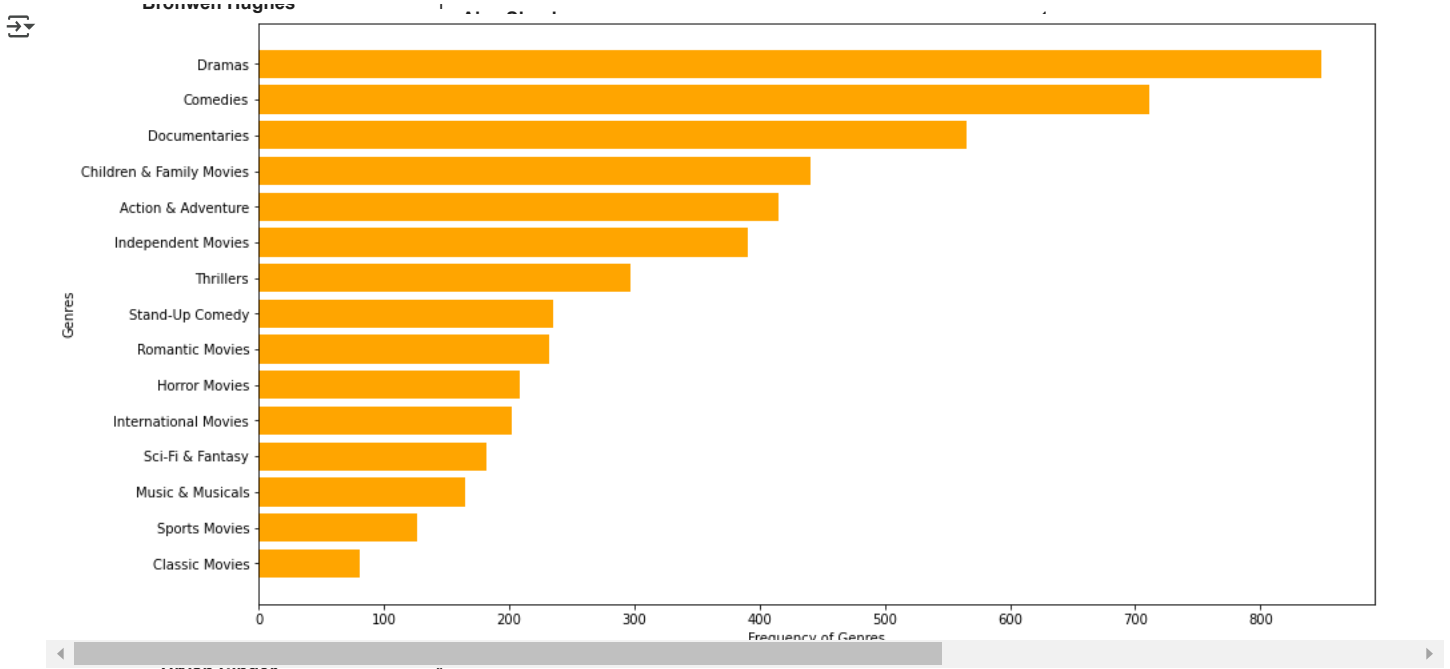
```
#Analyzing USA for both shows and movies
df_usa_shows=df_final1[df_final1['country']=='United States'][df_final1[df_final1['country']=='United States']['type']=='TV Show']
df_usa_movies=df_final1[df_final1['country']=='United States'][df_final1[df_final1['country']=='United States']['type']=='Movie']
```

```
df_genre=df_usa_shows.groupby(['Genre']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_genre[::-1]['Genre'], df_genre[::-1]['title'],color=['orange'])
plt.xlabel('Frequency of Genres')
plt.ylabel('Genres')
plt.show()
```



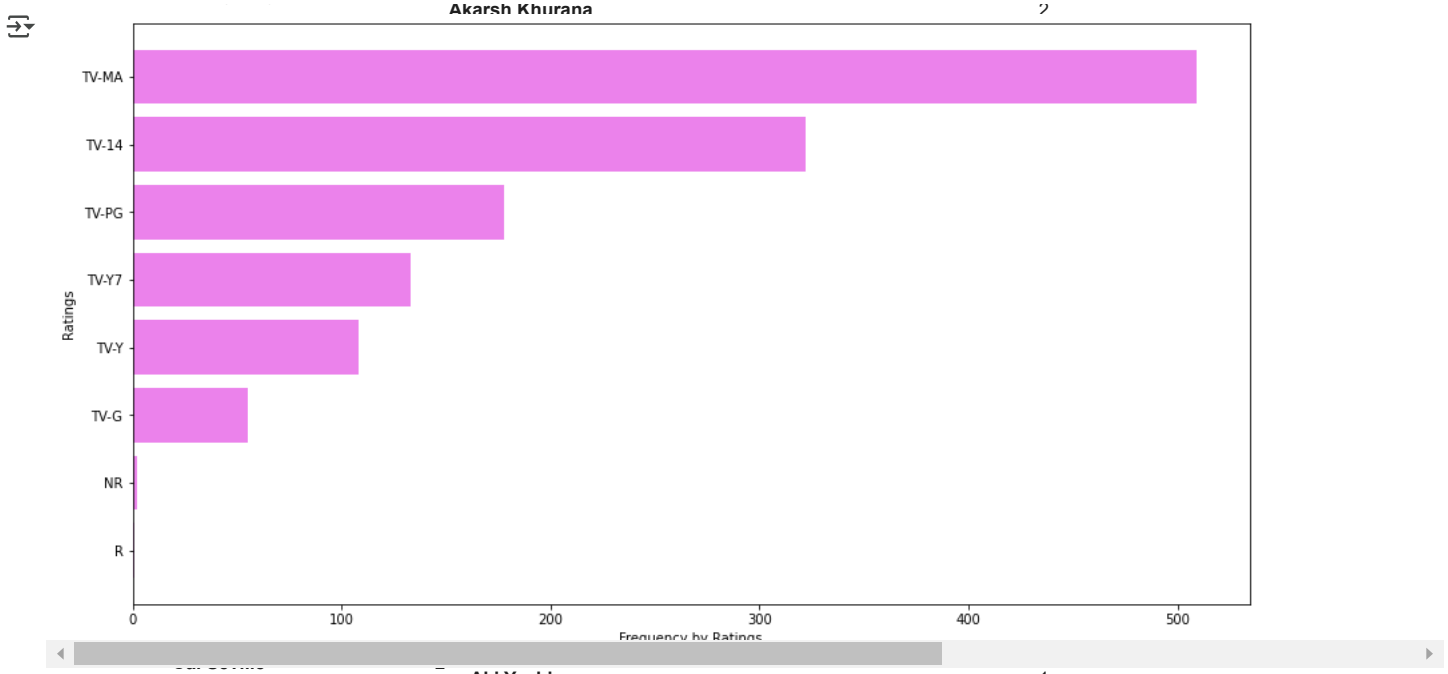
Dramas,Comedy, Kids 'TV Shows, International TV Shows and Docuseries, Genres are popular in TV Series in USA

```
df_genre=df_usa_movies.groupby(['Genre']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_genre[::-1]['Genre'], df_genre[::-1]['title'],color=['orange'])
plt.xlabel('Frequency of Genres')
plt.ylabel('Genres')
plt.show()
```

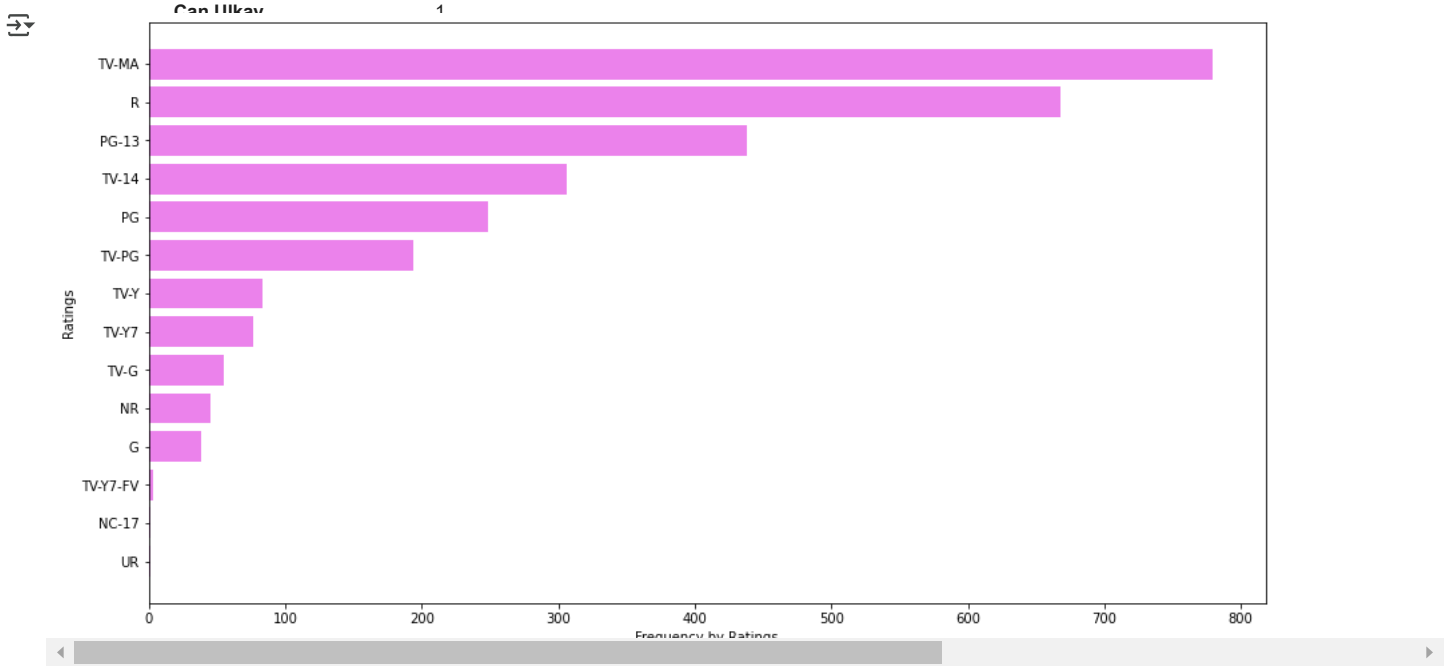


Dramas,Comedy, Documentaries, Family Movies and Action Genres in Movies are popular in USA

```
df_rating=df_usa_shows.groupby(['rating']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_rating[::-1]['rating'], df_rating[::-1]['title'],color=['violet'])
plt.xlabel('Frequency by Ratings')
plt.ylabel('Ratings')
plt.show()
```



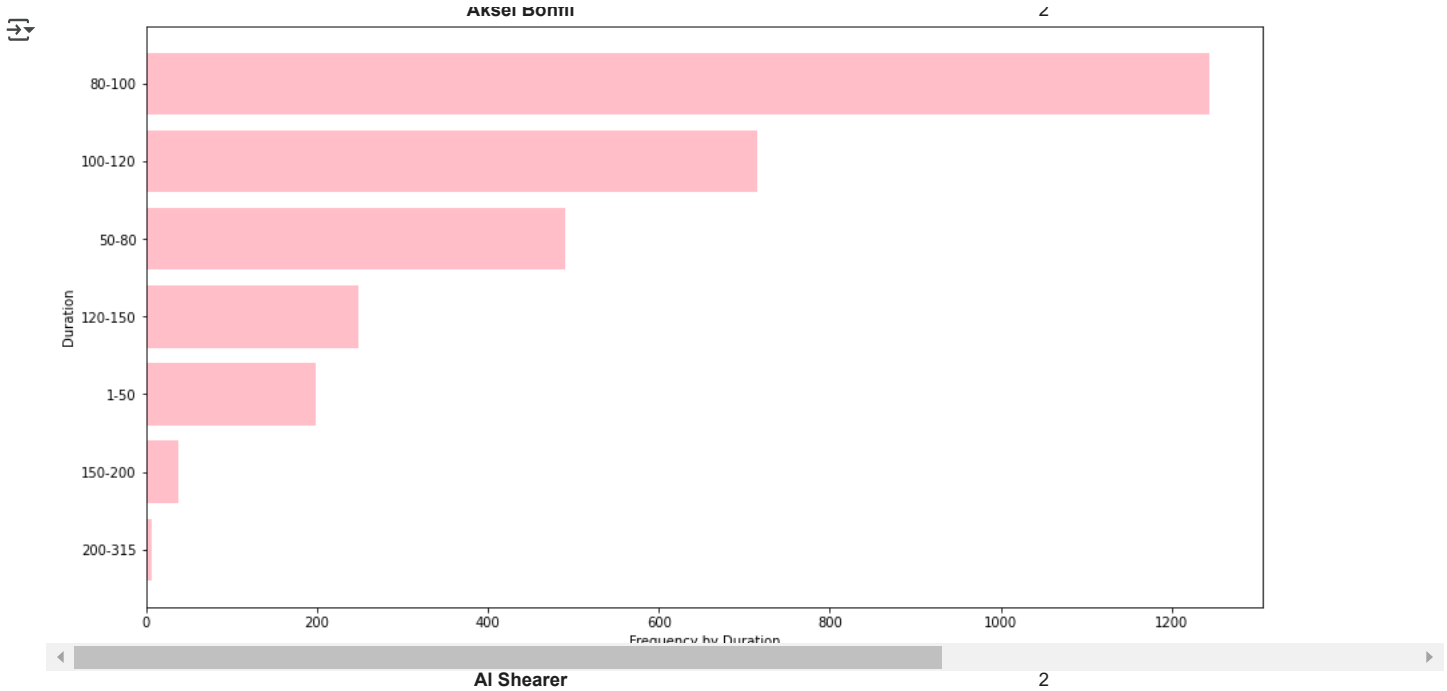

```
df_rating=df_usa_movies.groupby(['rating']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_rating[::-1]['rating'], df_rating[::-1]['title'],color=['violet'])
plt.xlabel('Frequency by Ratings')
plt.ylabel('Ratings')
plt.show()
```



So it seems plausible to conclude that the popular ratings across Netflix includes Mature Audiences and those appropriate for over 14/over 17 ages in both Movies and TV Shows in USA

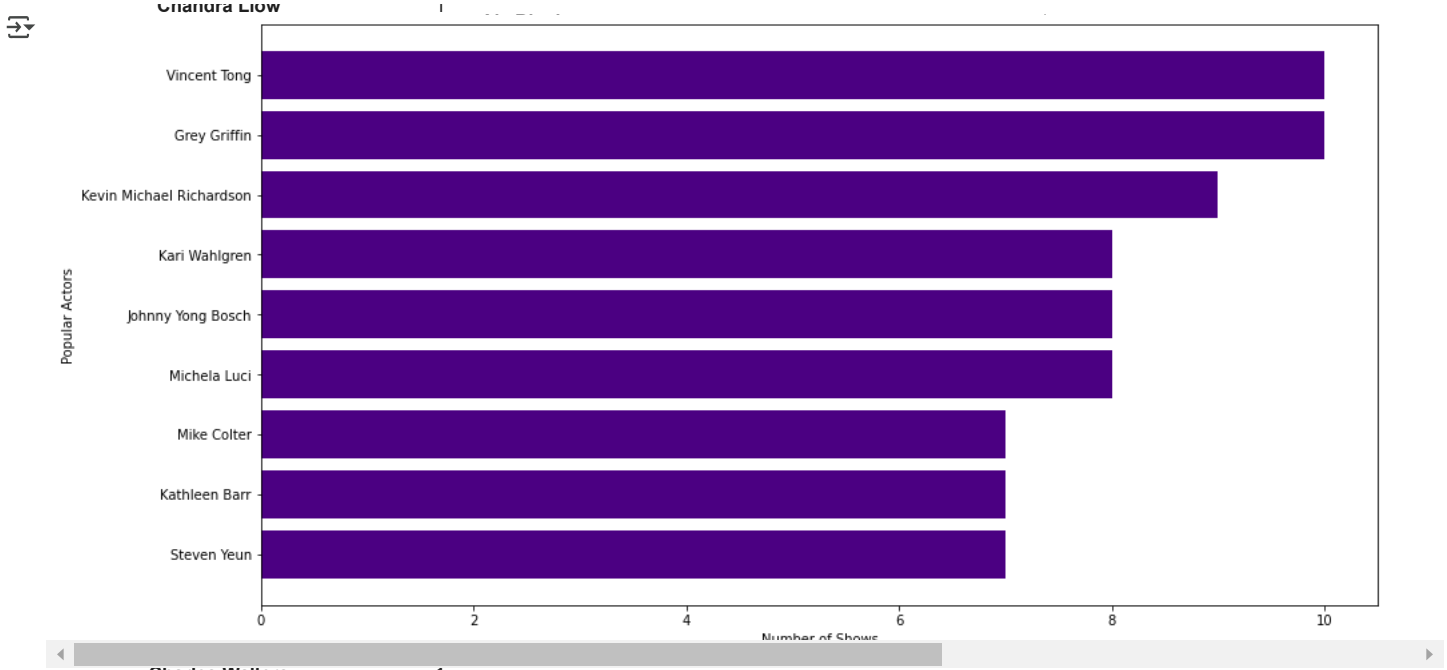
```
df_duration=df_usa_movies.groupby(['duration']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
plt.figure(figsize=(15,8))
plt.barh(df_duration[::-1]['duration'], df_duration[::-1]['title'],color=['pink'])
plt.xlabel('Frequency by Duration')
```

```
plt.ylabel('Duration')
plt.show()
```



Across movies 80-100,100-120 is the ranges of minutes for which most movies lie. So quite possibly 80-120 mins is the sweet spot we would be wanting for movies in USA

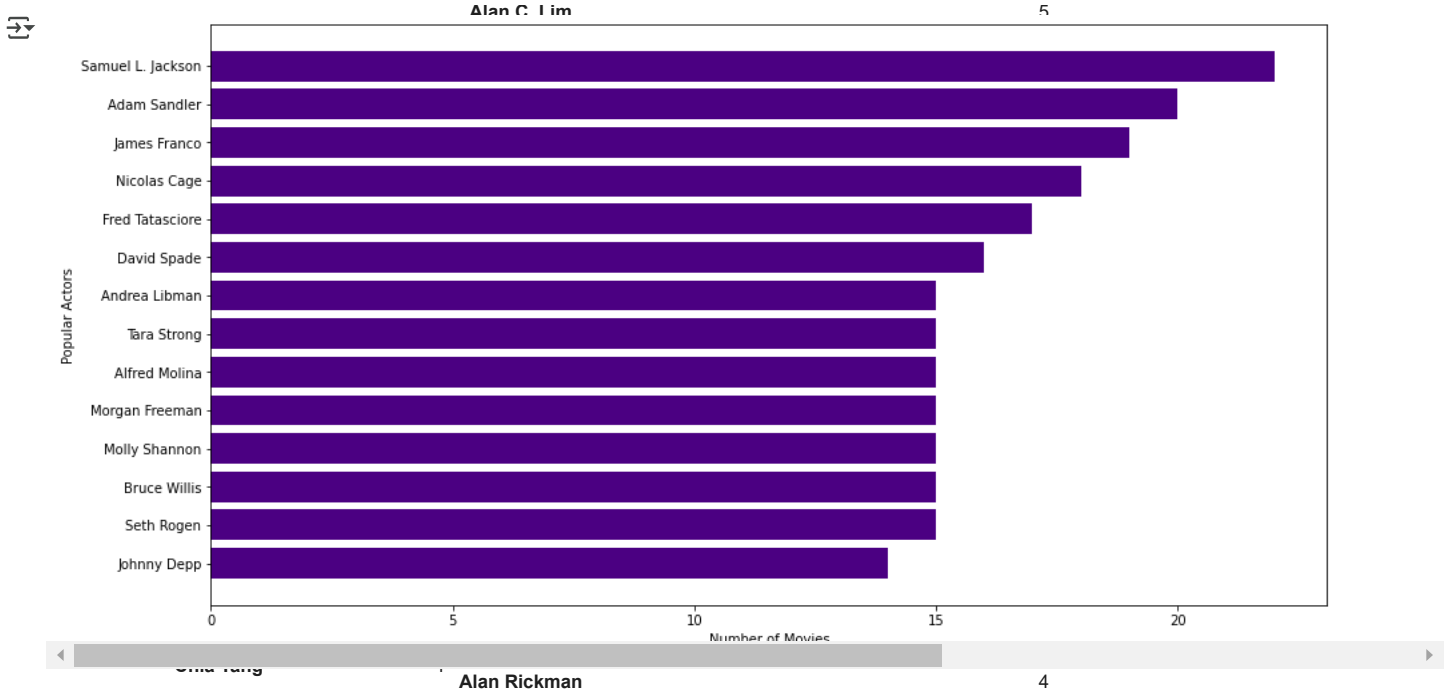
```
df_actors=df_usa_shows.groupby(['Actors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
df_actors=df_actors[df_actors['Actors']!='Unknown Actor']
plt.figure(figsize=(15,8))
plt.barh(df_actors[::-1]['Actors'], df_actors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Shows')
plt.ylabel('Popular Actors')
plt.show()
```



Vincent Tong,Grey Griffin and Kevin Richardson are the most popular actors across TV Shows in USA

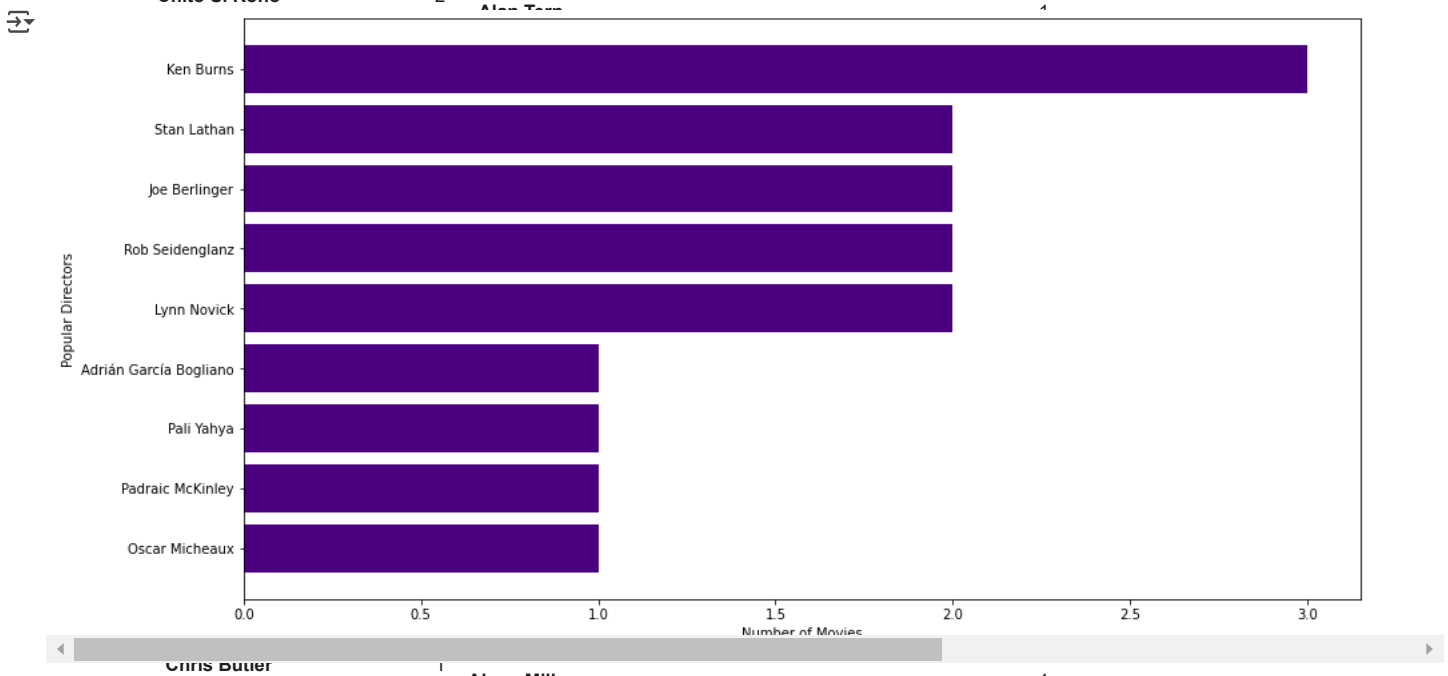
```
df_actors=df_usa_movies.groupby(['Actors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
df_actors=df_actors[df_actors['Actors']!='Unknown Actor']
plt.figure(figsize=(15,8))
plt.barh(df_actors[::-1]['Actors'], df_actors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
```

```
plt.ylabel('Popular Actors')
plt.show()
```



Samuel Jackson,Adam Sandler,James Franco and Nicolas Cage are very much popular across movies on Netflix in USA

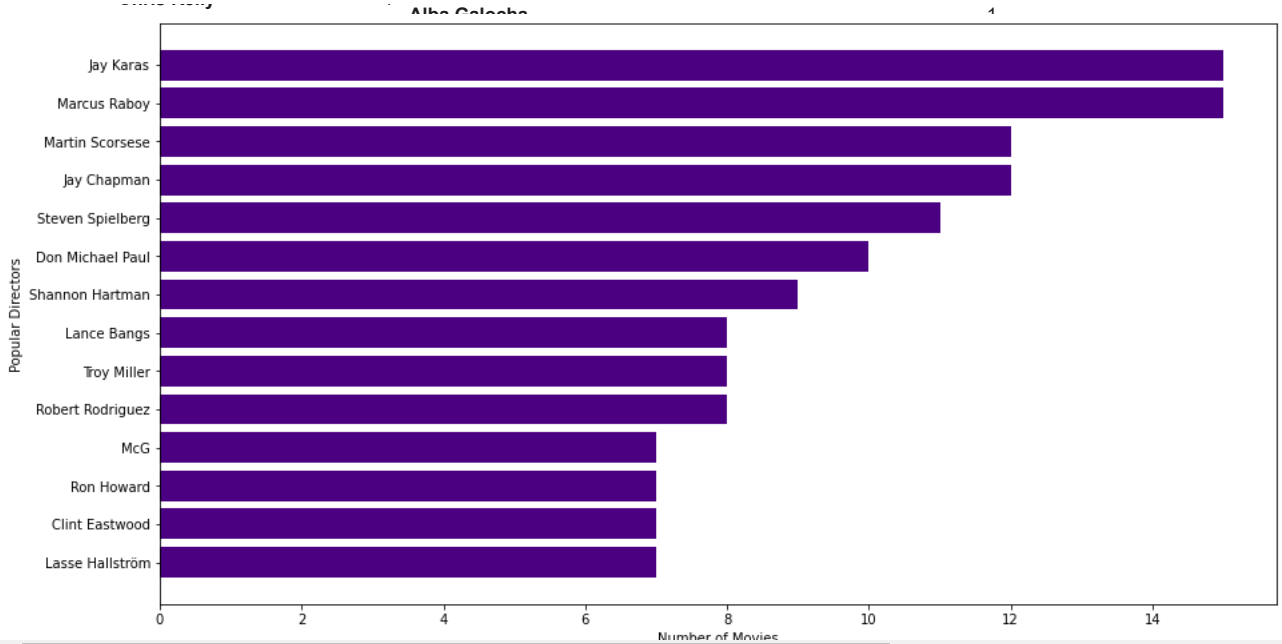
```
df_directors=df_usa_shows.groupby(['Directors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
df_directors=df_directors[df_directors['Directors']!='Unknown Director']
plt.figure(figsize=(15,8))
plt.barh(df_directors[::-1]['Directors'], df_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Directors')
plt.show()
```



Ken Burns,Stan Lathan,Joe Berlinger are popular directors across TV Shows on Netflix in USA

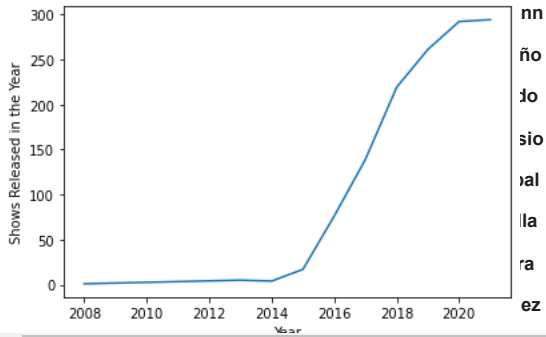
```
df_directors=df_usa_movies.groupby(['Directors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
df_directors=df_directors[df_directors['Directors']!='Unknown Director']
plt.figure(figsize=(15,8))
plt.barh(df_directors[::-1]['Directors'], df_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
```

```
plt.ylabel('Popular Directors')
plt.show()
```

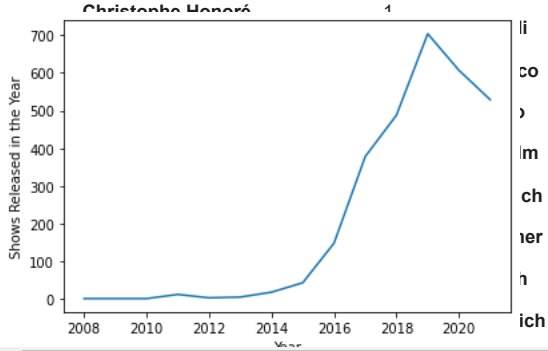


Jay Karas,Marcus Raboy,Martin Scorcese and Jay Chapman are popular directors across movies in USA

```
df_year=df_usa_shows.groupby(['year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_year, x='year', y='title')
plt.ylabel("Shows Released in the Year")
plt.xlabel("Year")
plt.show()
```

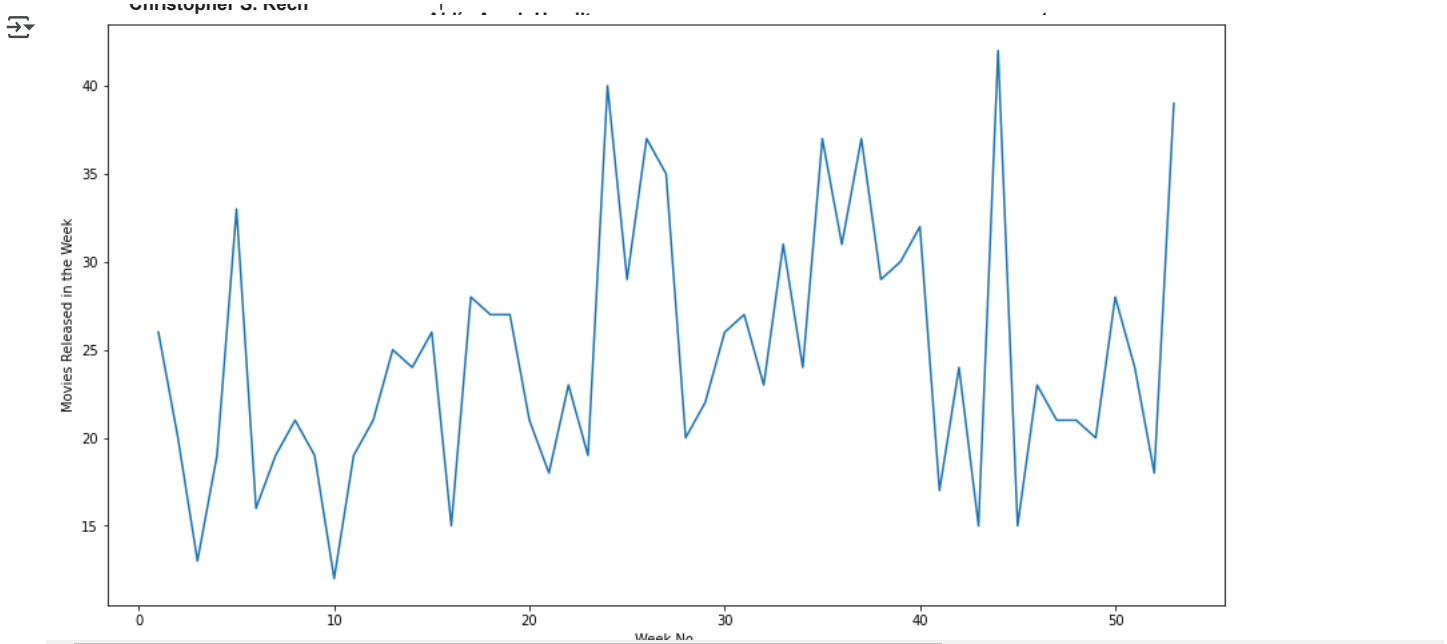



```
df_year=df_usa_movies.groupby(['year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_year, x='year', y='title')
plt.ylabel("Shows Released in the Year")
plt.xlabel("Year")
plt.show()
```

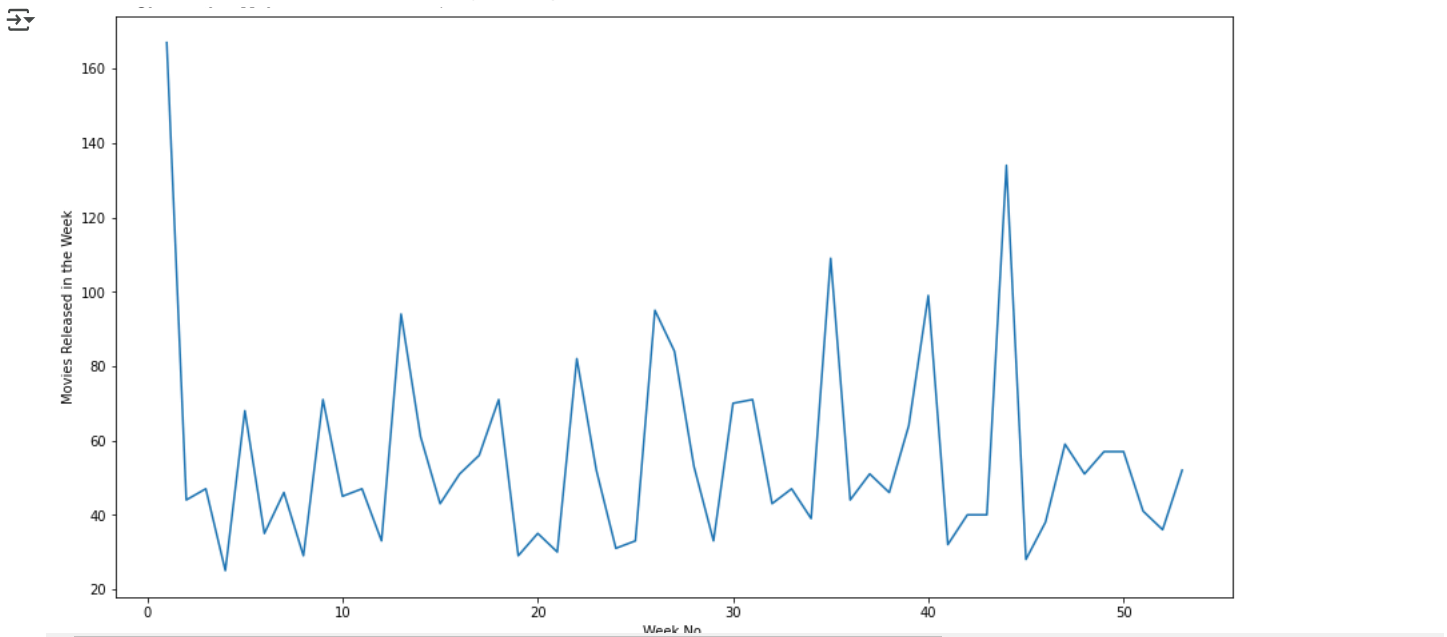


In USA, number of shows remained the same in 2021 as they were in 2020 while number of movies declined:

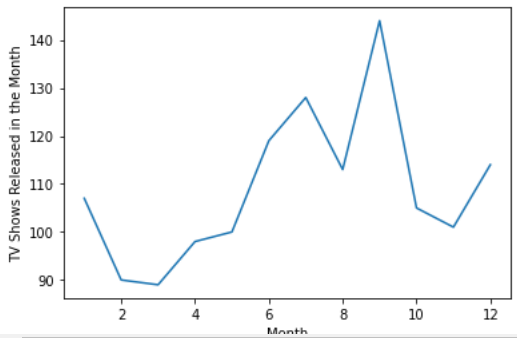
```
df_week=df_usa_shows.groupby(['week_Added']).agg({"title":"nunique"}).reset_index()
plt.figure(figsize=(15,8))
sns.lineplot(data=df_week, x='week_Added', y='title')
plt.ylabel("Movies Released in the Week")
plt.xlabel("Week No.")
plt.show()
```



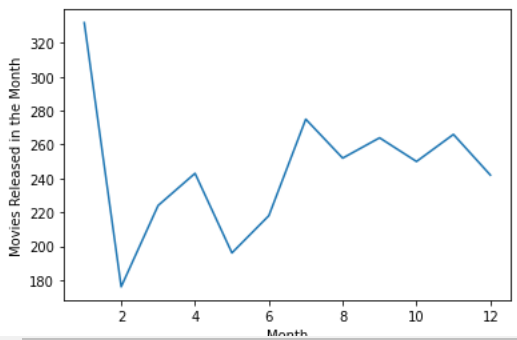

```
df_week=df_usa_movies.groupby(['week_Added']).agg({"title":"nunique"}).reset_index()
plt.figure(figsize=(15,8))
sns.lineplot(data=df_week, x='week_Added', y='title')
plt.ylabel("Movies Released in the Week")
plt.xlabel("Week No.")
plt.show()
```




```
df_month=df_usa_shows.groupby(['month_added']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_month, x='month_added', y='title')
plt.ylabel("TV Shows Released in the Month")
plt.xlabel("Month")
plt.show()
```



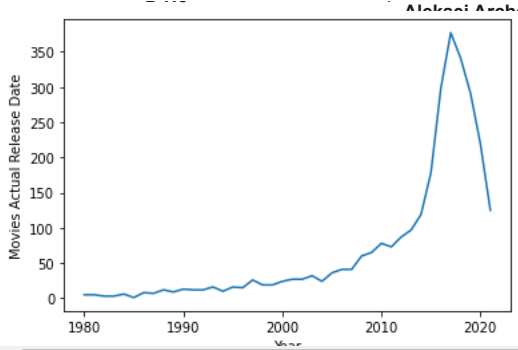

```
df_month=df_usa_movies.groupby(['month_added']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_month, x='month_added', y='title')
plt.ylabel("Movies Released in the Month")
plt.xlabel("Month")
plt.show()
```



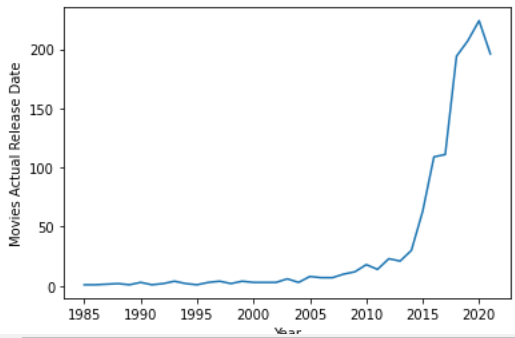
TV Shows are added in Netflix by a tremendous amount in July and September in USA

Movies are added in Netflix in USA by a tremendous amount in first week/last month of current year and first month of next year

```
df_release_year=df_usa_movies[df_usa_movies['release_year']>=1980].groupby(['release_year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_release_year, x='release_year', y='title')
plt.ylabel("Movies Actual Release Date")
plt.xlabel("Year")
plt.show()
```




```
df_release_year=df_usa_shows[df_usa_shows['release_year']>=1980].groupby(['release_year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_release_year, x='release_year', y='title')
plt.ylabel("Movies Actual Release Date")
plt.xlabel("Year")
plt.show()
```



In USA, though both Movies and Shows have reduced in 2021, the amount of decrease in number of TV Shows is small as compared to Movies

```
df_usa_movies.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	Modified_Added_da
0	Dick Johnson Is Dead	Unknown Actor	Kirsten Johnson	Documentaries	United States	s1	Movie	September 25, 2021	2020	PG-13	80-100	2021-09-
159	My Little Pony: A New Generation	Vanessa Hudgens	Robert Cullen	Children & Family Movies	United States	s7	Movie	September 24, 2021	2021	PG	80-100	2021-09-
160	My Little Pony: A New Generation	Vanessa Hudgens	José Luis Ucha	Children & Family Movies	United States	s7	Movie	September 24, 2021	2021	PG	80-100	2021-09-

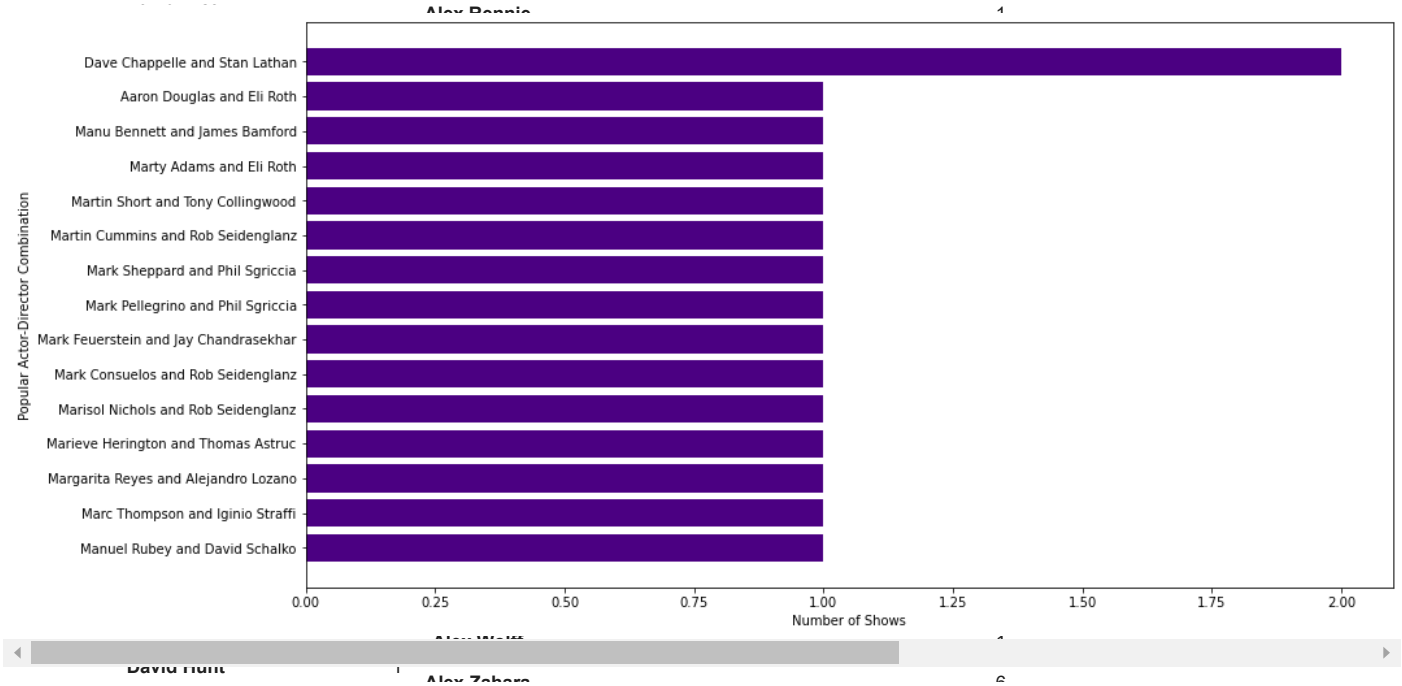
```
#Analysing a combination of actors and directors
df_usa_movies['Actor_Director_Combination'] = df_usa_movies.Actors.str.cat(df_usa_movies.Directors, sep=' and ')
df_usa_movies_subset=df_usa_movies[df_usa_movies['Actors']!='Unknown Actor']
df_usa_movies_subset=df_usa_movies_subset[df_usa_movies_subset['Directors']!='Unknown Director']
df_usa_movies_subset.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	Modified_Added_date	mo
159	My Little Pony: A New Generation	Vanessa Hudgens	Robert Cullen	Children & Family Movies	United States	s7	Movie	September 24, 2021	2021	PG	80-100	2021-09-24	
160	My Little Pony: A New Generation	Vanessa Hudgens	José Luis Ucha	Children & Family Movies	United States	s7	Movie	September 24, 2021	2021	PG	80-100	2021-09-24	
161	My Little Pony: A New Generation	Kimiko Glenn	Robert Cullen	Children & Family Movies	United States	s7	Movie	September 24, 2021	2021	PG	80-100	2021-09-24	
162	My Little Pony: A New Generation	Kimiko Glenn	José Luis Ucha	Children & Family Movies	United States	s7	Movie	September 24, 2021	2021	PG	80-100	2021-09-24	
163	My Little Pony: A New Generation	James Marsden	Robert Cullen	Children & Family Movies	United States	s7	Movie	September 24, 2021	2021	PG	80-100	2021-09-24	

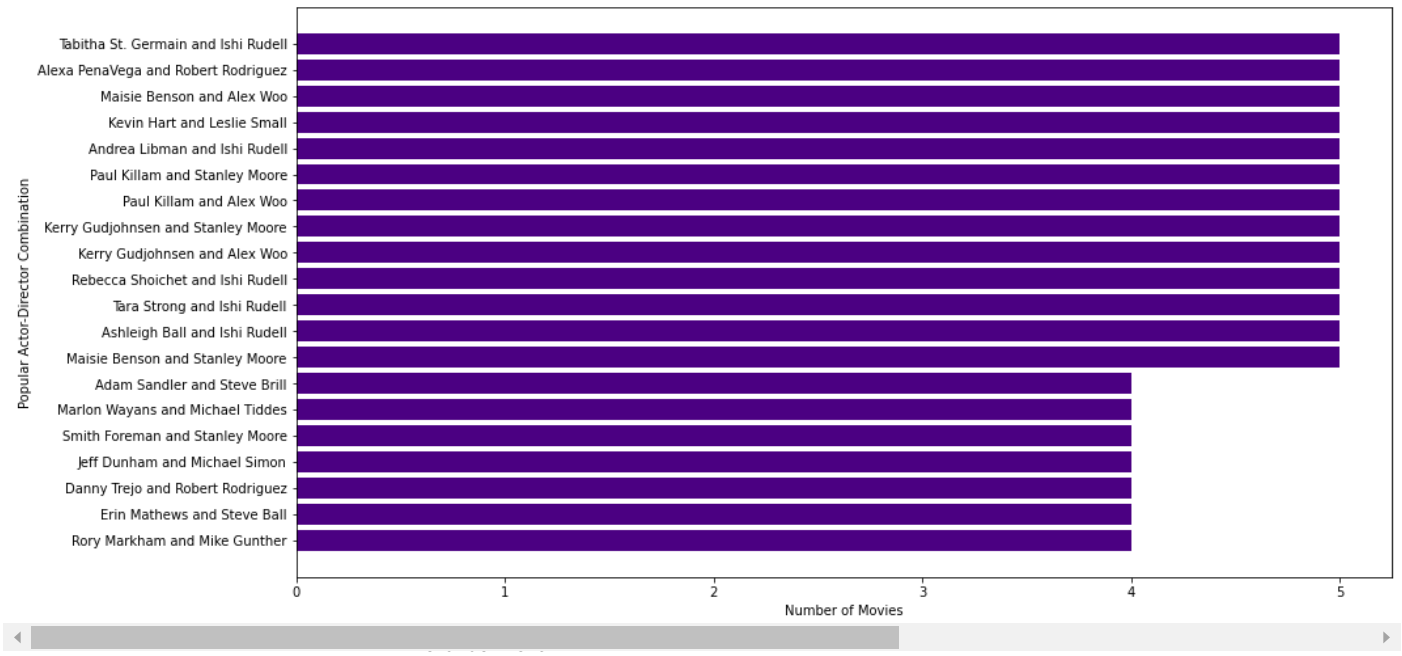
```
df_usa_shows['Actor_Director_Combination'] = df_usa_shows.Actors.str.cat(df_usa_shows.Directors, sep=' and ')
df_usa_shows_subset=df_usa_shows[df_usa_shows['Actors']!='Unknown Actor']
df_usa_shows_subset=df_usa_shows_subset[df_usa_shows_subset['Directors']!='Unknown Director']
df_usa_shows_subset.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	Modified_Added_date	month
111	Midnight Mass	Kate Siegel	Mike Flanagan	TV Dramas	United States	s6	TV Show	September 24, 2021	2021	TV-MA	1 Season	2021-09-24	
112	Midnight Mass	Kate Siegel	Mike Flanagan	TV Horror	United States	s6	TV Show	September 24, 2021	2021	TV-MA	1 Season	2021-09-24	
113	Midnight Mass	Kate Siegel	Mike Flanagan	TV Mysteries	United States	s6	TV Show	September 24, 2021	2021	TV-MA	1 Season	2021-09-24	
114	Midnight Mass	Zach Gilford	Mike Flanagan	TV Dramas	United States	s6	TV Show	September 24, 2021	2021	TV-MA	1 Season	2021-09-24	
115	Midnight Mass	Zach Gilford	Mike Flanagan	TV Horror	United States	s6	TV Show	September 24, 2021	2021	TV-MA	1 Season	2021-09-24	

```
df_actors_directors=df_usa_shows_subset.groupby(['Actor_Director_Combination']).agg({"title":"nunique"}).reset_index().sort_values(by=['titl
plt.figure(figsize=(15,8))
plt.barh(df_actors_directors[::-1]['Actor_Director_Combination'], df_actors_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Shows')
plt.ylabel('Popular Actor-Director Combination')
plt.show()
```




```
df_actors_directors=df_usa_movies_subset.groupby(['Actor_Director_Combination']).agg({"title":"nunique"}).reset_index().sort_values(by=['tit
plt.figure(figsize=(15,8))
plt.barh(df_actors_directors[::-1]['Actor_Director_Combination'], df_actors_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Actor-Director Combination')
plt.show()
```




```
df_actors_directors[::-1]['Actor_Director_Combination'].values
```

```
array(['Rory Markham and Mike Gunther', 'Erin Mathews and Steve Ball',
       'Danny Trejo and Robert Rodriguez',
       'Jeff Dunham and Michael Simon', 'Smith Foreman and Stanley Moore',
       'Marlon Wayans and Michael Tiddes', 'Adam Sandler and Steve Brill',
       'Maisie Benson and Stanley Moore', 'Ashleigh Ball and Ishi Rudell',
       'Tara Strong and Ishi Rudell', 'Rebecca Shoichet and Ishi Rudell',
       'Kerry Gudjohnsen and Alex Woo',
       'Kerry Gudjohnsen and Stanley Moore', 'Paul Killam and Alex Woo',
       'Paul Killam and Stanley Moore', 'Andrea Libman and Ishi Rudell',
       'Kevin Hart and Leslie Small', 'Maisie Benson and Alex Woo',
       'Alexa PenaVega and Robert Rodriguez',
       'Tabitha St. Germain and Ishi Rudell'], dtype=object)
```

The Most Popular Actor Director Combination in Movies Across USA are:-

'Smith Foreman and Stanley Moore',
'Marlon Wayans and Michael Tiddes',
'Adam Sandler and Steve Ball',
'Maisie Benson and Stanley Moore',
'Ashleigh Ball and Ishi Rudell',
'Tara Strong and Ishi Rudell',
'Rebecca Shoichet and Ishi Rudell',
'Kerry Gudjohnsen and Alex Woo',
'Kerry Gudjohnsen and Stanley Moore',
'Paul Killam and Alex Woo',
'Paul Killam and Stanley Moore',
'Andrea Libman and Ishi Rudell',
'Kevin Hart and Leslie Small',
'Maisie Benson and Alex Woo',
'Alexa PenaVega and Robert Rodriguez',
'Tabitha St. Germain and Ishi Rudell'

The Second Most Popular Actor Director Combination in Movies Across USA are:-

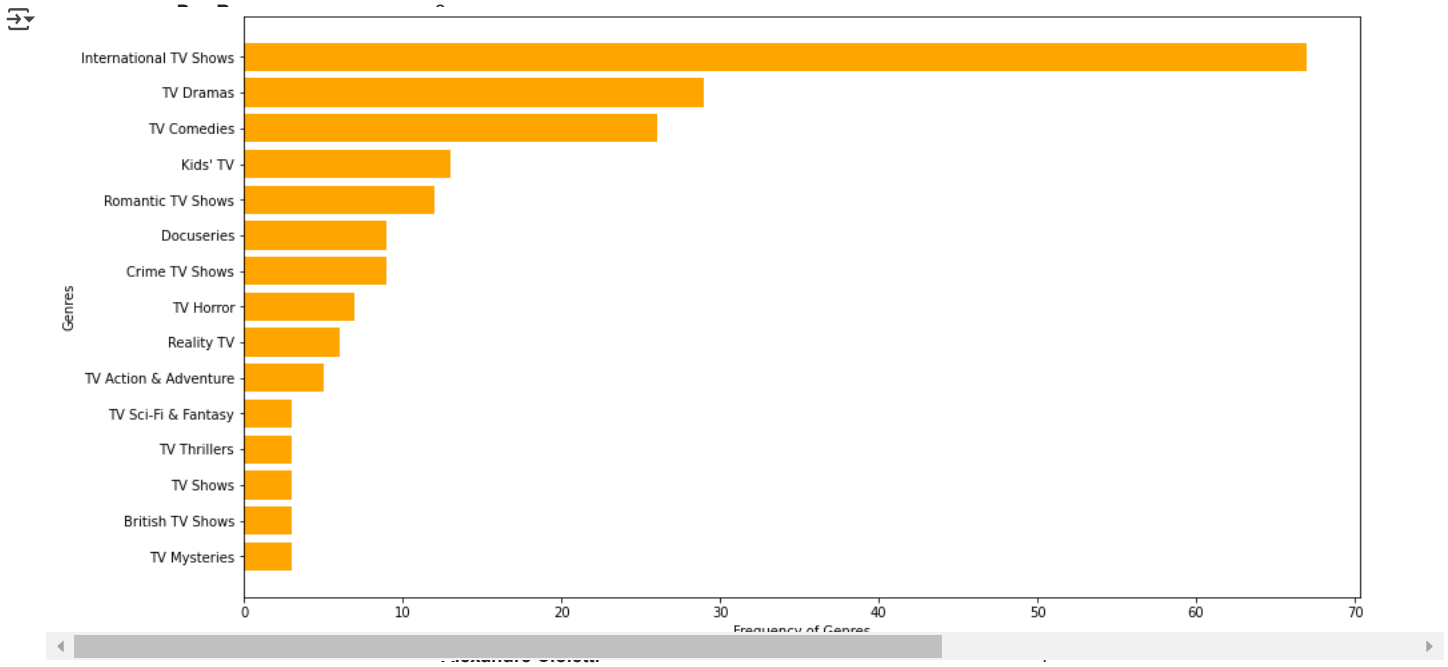
'Rory Markham and Mike Gunther',
'Erin Mathews and Steve Ball',
'Danny Trejo and Robert Rodriguez',
'Jeff Dunham and Michael Simon'

Alexander Holtmann	2
Alexander Koch	1
Alexander Lloyd	1
lexander Ludwig	3
Alexander Martin	1
lexander Morton	1
Alexander Oblasov	1
lexander Peleg	1
Alexander Petrov	3
lexander Scheer	1
Alexander Siddig	6
xander Skarsgård	4
Alexander Strobele	1
lexander Tsekalo	1
Alexander Vlahos	1
lexander Wraith	1
Alexander Yatsenko	1
lexander Zwart	1
Alexandra Ayo	1
lexandra Beaton	1
Alexandra Borbély	1
Alexandre Goyette	1
lexandre Harough	1

Univariate Analysis separately for shows and movies in India

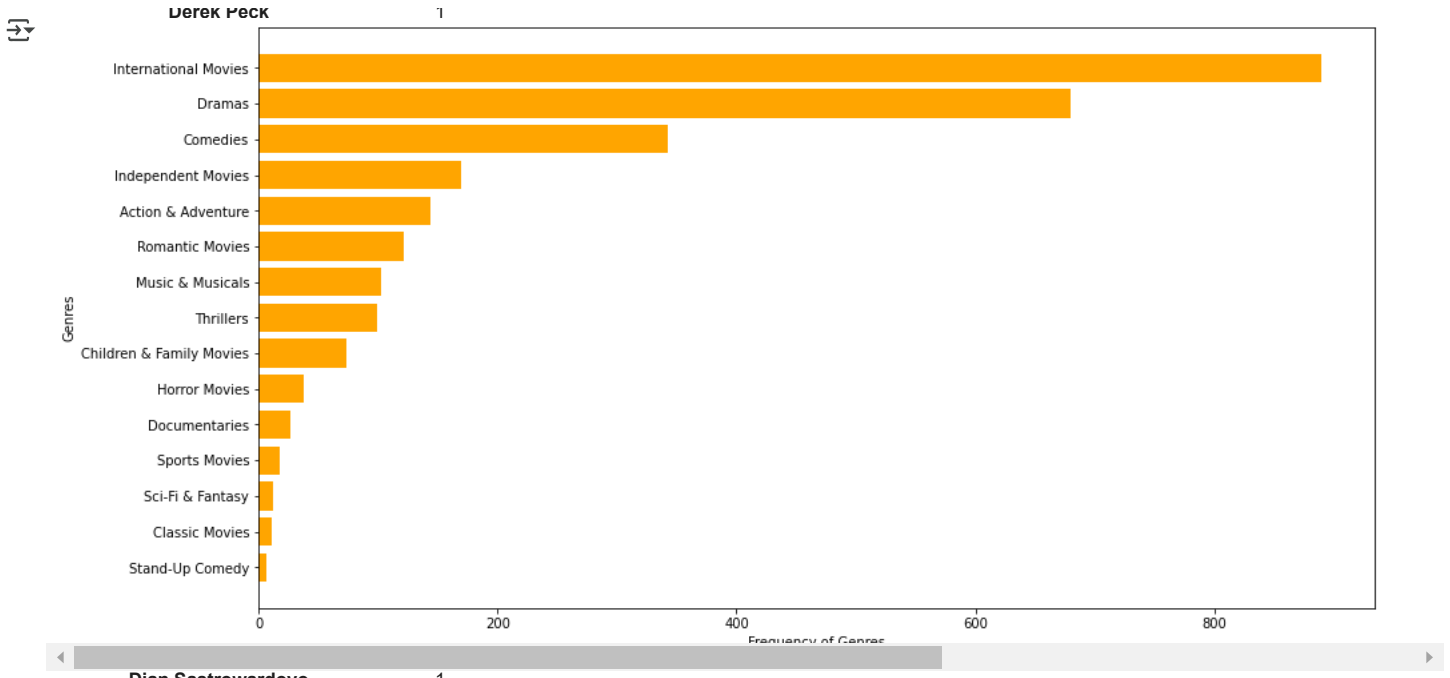
```
#Analyzing India for both shows and movies
df_india_shows=df_final1[df_final1['country']=='India'][df_final1[df_final1['country']=='India']['type']=='TV Show']
df_india_movies=df_final1[df_final1['country']=='India'][df_final1[df_final1['country']=='India']['type']=='Movie']
```

```
df_genre=df_india_shows.groupby(['Genre']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_genre[::-1]['Genre'], df_genre[::-1]['title'],color=['orange'])
plt.xlabel('Frequency of Genres')
plt.ylabel('Genres')
plt.show()
```



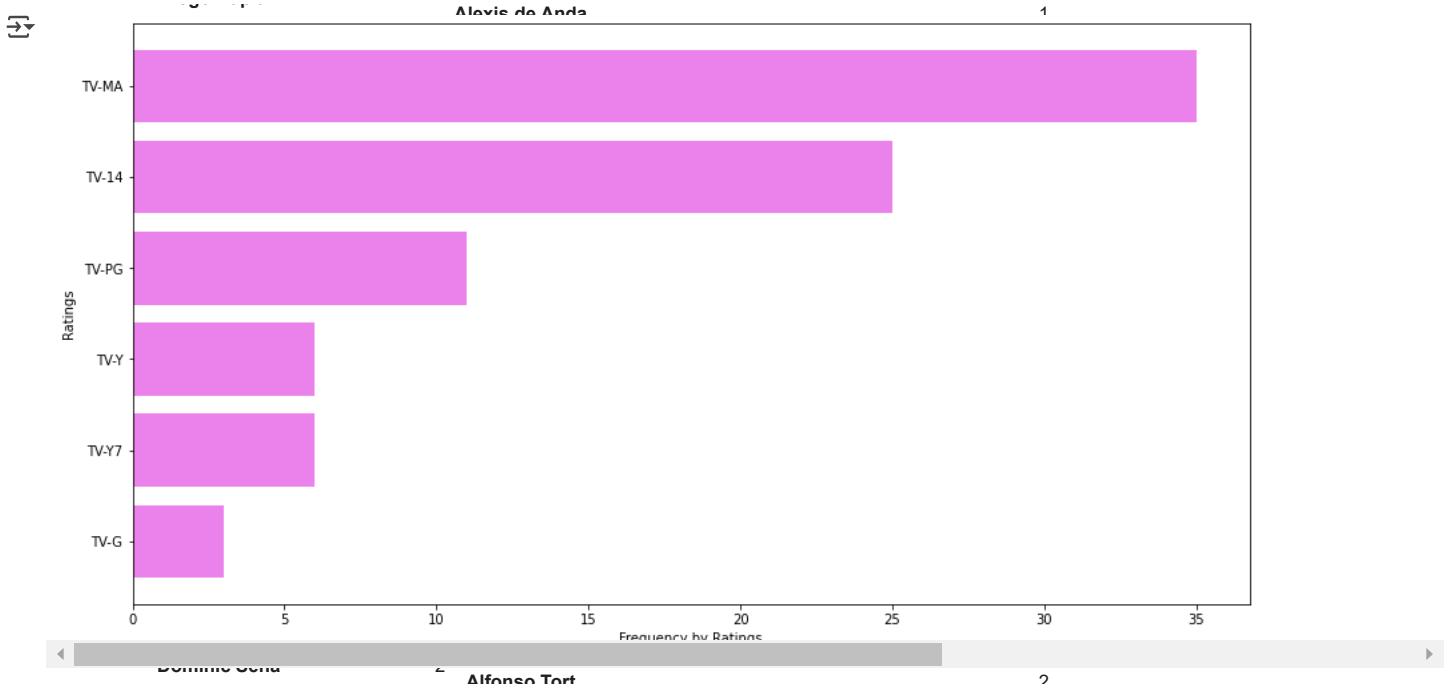
Dramas,Comedy, Kids 'TV Shows and International TV Shows Genres are popular in TV Series in India

```
df_genre=df_india_movies.groupby(['Genre']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_genre[::-1]['Genre'], df_genre[::-1]['title'],color=['orange'])
plt.xlabel('Frequency of Genres')
plt.ylabel('Genres')
plt.show()
```

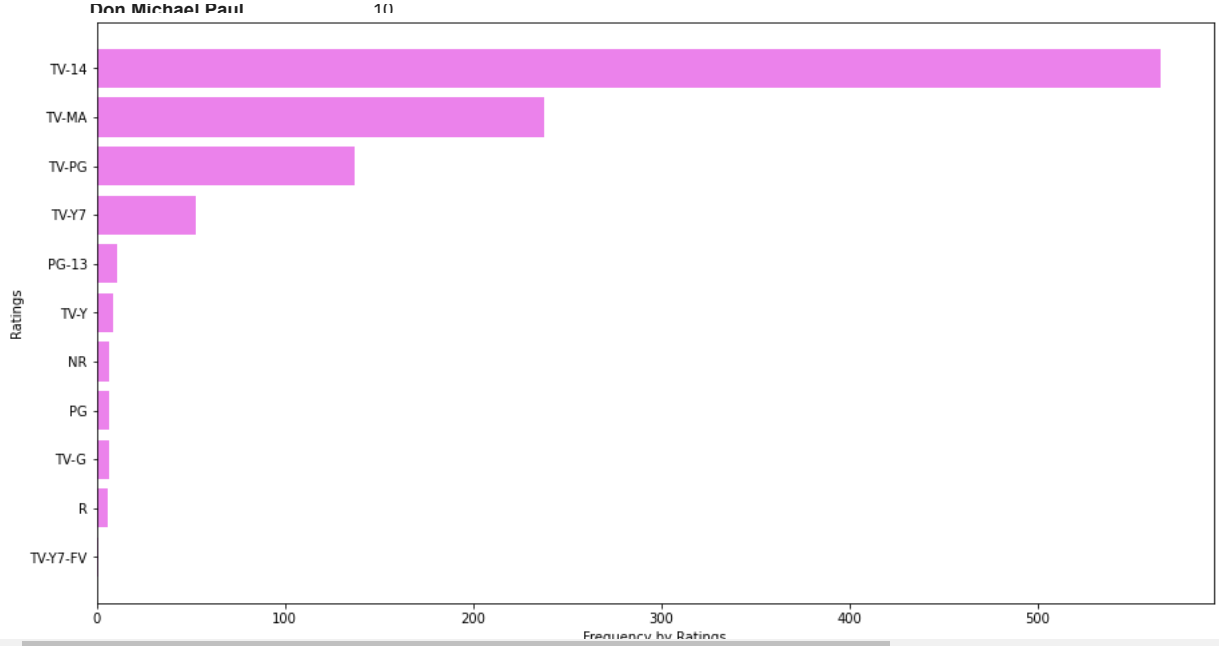


International Movies Drama Comedy,Indpeendent Movies and Action, Romance Genres are prevalent in India

```
df_rating=df_india_shows.groupby(['rating']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_rating[::-1]['rating'], df_rating[::-1]['title'],color=['violet'])
plt.xlabel('Frequency by Ratings')
plt.ylabel('Ratings')
plt.show()
```



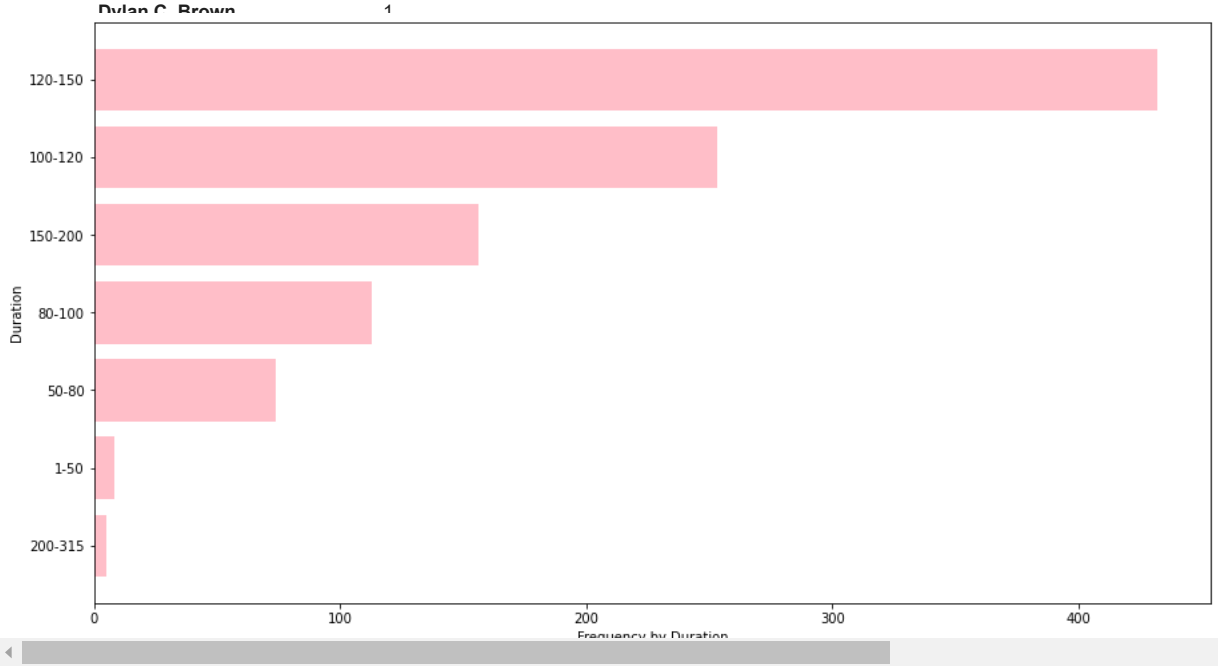

```
df_rating=df_india_movies.groupby(['rating']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_rating[::-1]['rating'], df_rating[::-1]['title'],color=['violet'])
plt.xlabel('Frequency by Ratings')
plt.ylabel('Ratings')
plt.show()
```



So it seems plausible to conclude that the popular ratings across Netflix includes Mature Audiences in TV Shows and those appropriate for people over 14 in Movies in India.

Now this indeed seems to be the case. Indian TV Shows in Netflix are without a shadow of doubt intended for Mature Audiences while Movies for over 14 years of age.

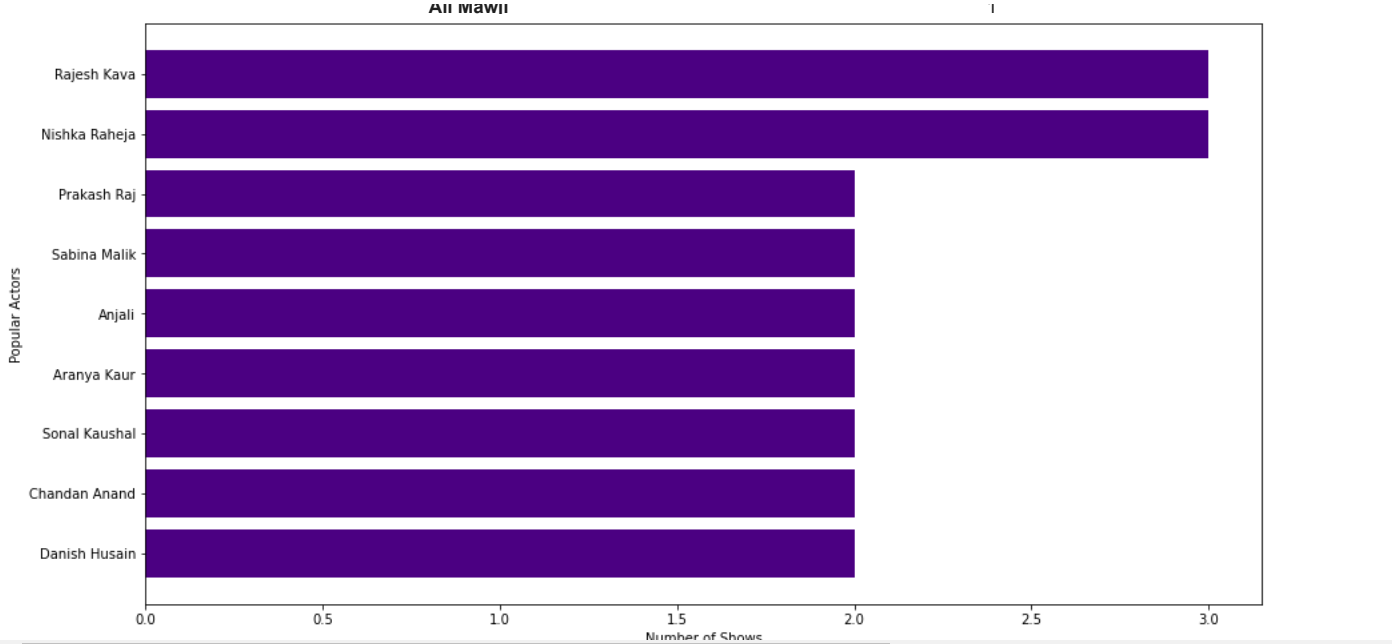
```
df_duration=df_india_movies.groupby(['duration']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
plt.figure(figsize=(15,8))
plt.barh(df_duration[::-1]['duration'], df_duration[::-1]['title'],color=['pink'])
plt.xlabel('Frequency by Duration')
plt.ylabel('Duration')
plt.show()
```



Across movies ranges of minutes in India are comparatively greater than USA with a sweet spot at 120-150 mins.

```
df_actors=df_india_shows.groupby(['Actors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
df_actors=df_actors[df_actors['Actors']!='Unknown Actor']
plt.figure(figsize=(15,8))
plt.barh(df_actors[::-1]['Actors'], df_actors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Shows')
```

```
plt.ylabel('Popular Actors')
plt.show()
```



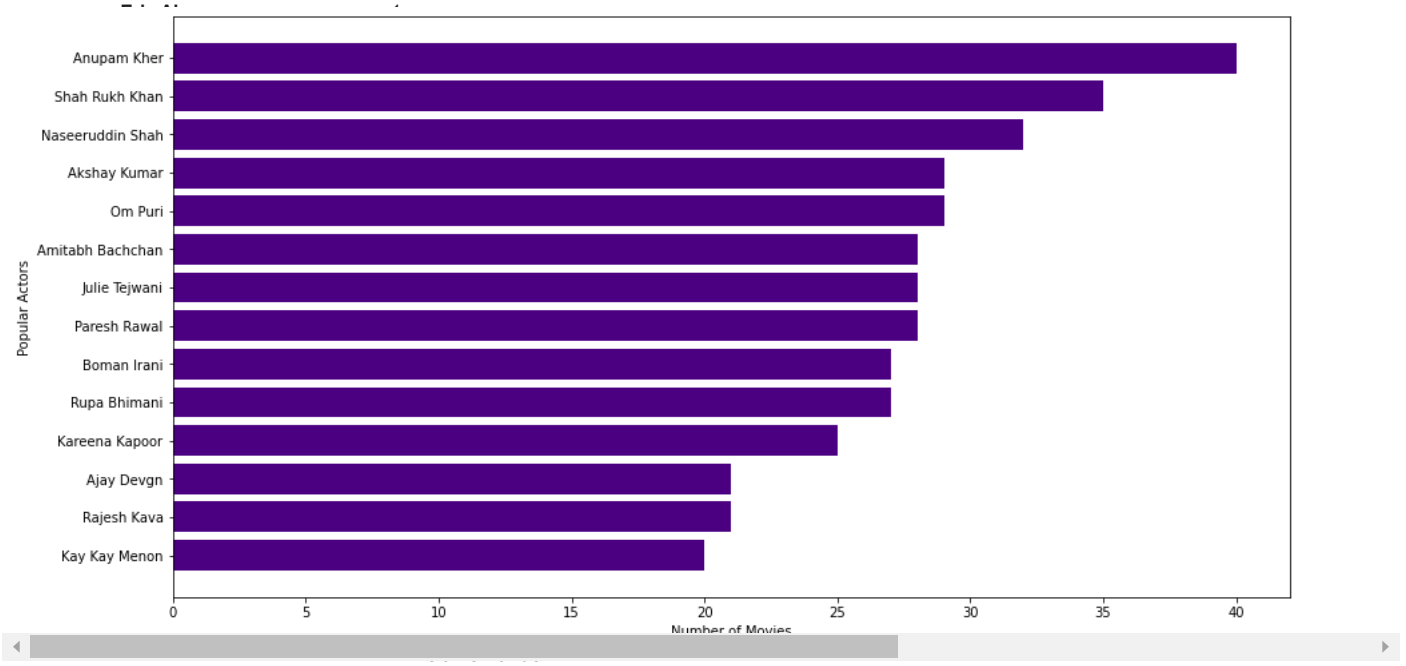

```
df_actors['Actors'].values
```

```
array(['Rajesh Kava', 'Nishka Raheja', 'Prakash Raj', 'Sabina Malik',
       'Anjali', 'Aranya Kaur', 'Sonal Kaushal', 'Chandan Anand',
       'Danish Husain'], dtype=object)
```

Popular Actors in TV Shows in India are:-

'Rajesh Kava',
'Nishka Raheja',
'Prakash Raj',
'Sabina Malik',
'Anjali',
'Aranya Kaur',
'Sonal Kaushal',
'Chandan Anand',
'Danish Husain'

```
df_actors=df_india_movies.groupby(['Actors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
df_actors=df_actors[df_actors['Actors']!='Unknown Actor']
plt.figure(figsize=(15,8))
plt.barh(df_actors[::-1]['Actors'], df_actors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Actors')
plt.show()
```




```
df_actors['Actors'].values
```

```
array(['Anupam Kher', 'Shah Rukh Khan', 'Naseeruddin Shah',
       'Akshay Kumar', 'Om Puri', 'Amitabh Bachchan', 'Julie Tejwani',
       'Paresh Rawal', 'Boman Irani', 'Rupa Bhimani', 'Kareena Kapoor',
       'Ajay Devgn', 'Rajesh Kava', 'Kay Kay Menon'], dtype=object)
```

Popular actors across Movies in India:-

'Anupam Kher',
'Shah Rukh Khan',
'Naseeruddin Shah',
'Akshay Kumar',
'Om Puri',
'Paresh Rawal',
'Julie Tejwani',
'Amitabh Bachchan',
'Boman Irani',
'Rupa Bhimani',
'Kareena Kapoor',
'Ajay Devgn',
'Rajesh Kava',
'Kay Kay Menon'

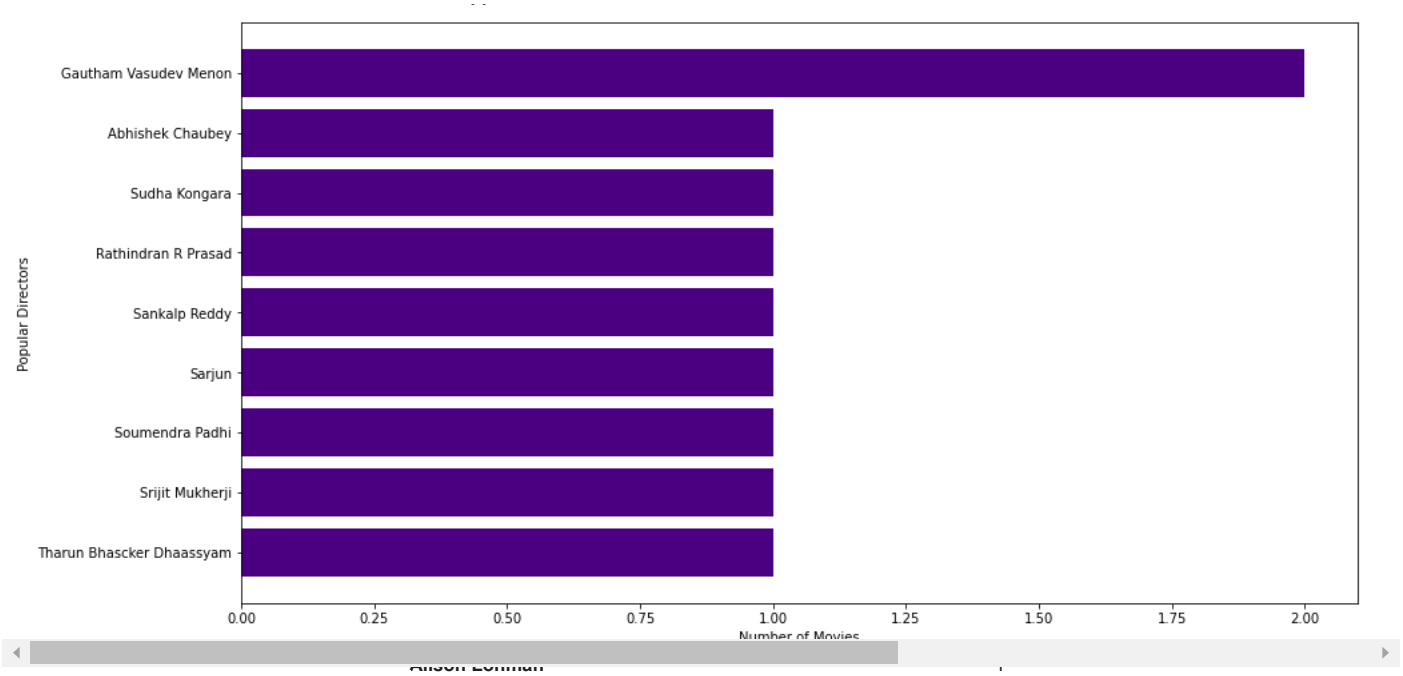
Eric Zaragosa	1
Erich Stürm	1
Erik Canuel	2
Erik Clemens	1
Erik Kling	1
Erik Matti	1
Erik Nelson	1
Erik White	1
Ernie Barbarash	4
Erol Özlevi	1
Errol Morris	2
Ertanto Robby Soediskam	1
Erwin van den Eshof	2
Eshom Nelms	1
Espen Sandberg	1
Esteban Crespo	2

Alicia James	1
Alicia Jaziz	1
Alicia Keys	1
Alicia Leigh Willis	1
Alicia Quiñonez	1
Alicia Rubio	2
Alicia Sandoval	1
Alicia Sanz	2
licia Silverstone	5
Alicia Sánchez	2
Alicia Vikander	7
Alicia Witt	1
icia von Rittberg	1
Alicyn Packard	3
la Baldari Calabria	1
Alie Ward	2
Alien Huang	1

```
df_directors=df_india_shows.groupby(['Directors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
df_directors=df_directors[df_directors['Directors']!='Unknown Director']
plt.figure(figsize=(15,8))
plt.barh(df_directors[::-1]['Directors'], df_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Directors')
plt.show()
```

Ethan Hawke	1
Eugenio Derbez	1
Eva Müller	1
Eva Orner	1
Eva Trobisch	1
Eva Vives	1
Evan Goldberg	1
Evan Katz	1

Alina Boz	1
Alina Cheng	1
Alina Chivulescu	1
Alina Foley	1
Alina Kukushkina	1
Alina Lozano	2
Alina Stiegler	1
Aline Borges	1
Aline Küppenheim	1




```
df_directors['Directors'].values
```

```
array(['Gautham Vasudev Menon', 'Abhishek Chaubey', 'Sudha Kongara',
       'Rathindran R Prasad', 'Sankalp Reddy', 'Sarjun',
       'Soumendra Padhi', 'Srijit Mukherji', 'Tharun Bhascker Dhaassyam'],
      dtype=object)
```

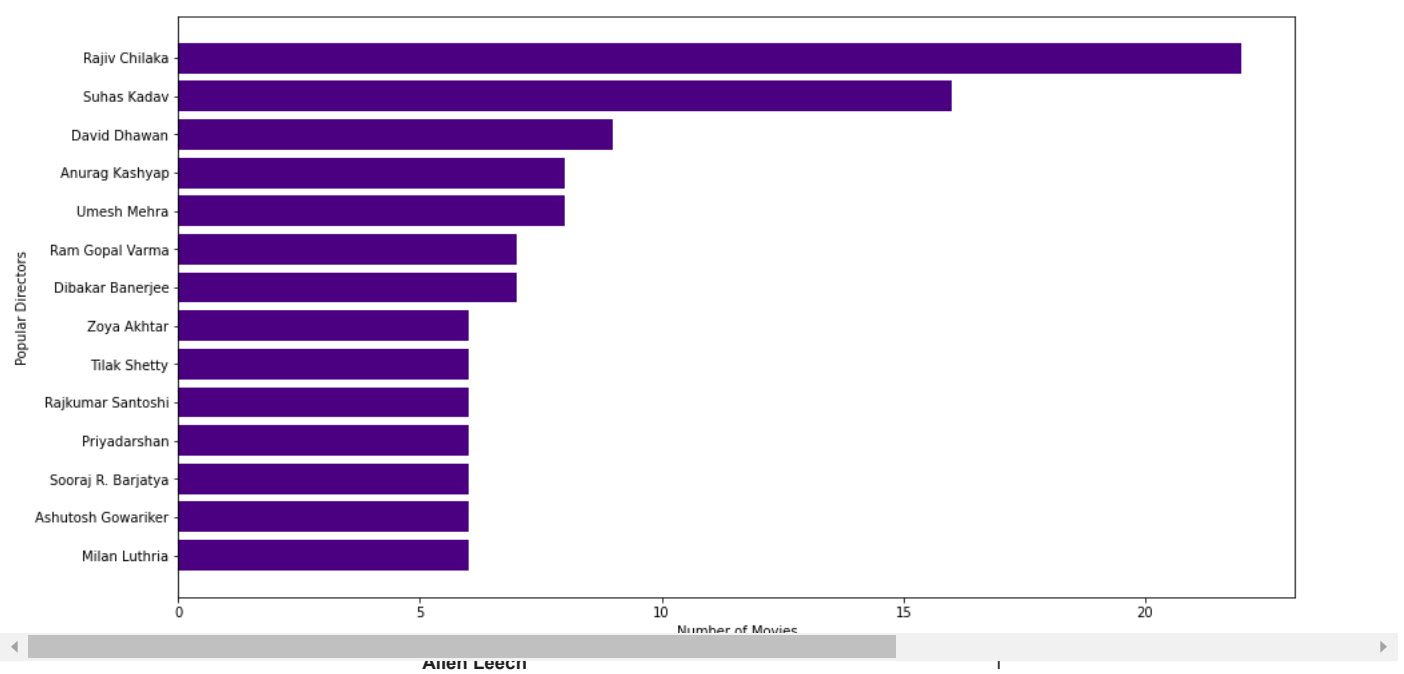
Popular Directors Across Movies in India:-

'Gautham Vasudev Menon',
'Abhishek Chaubey',
'Sudha Kongara',
'Rathindran R Prasad',
'Sankalp Reddy',
'Sarjun',
'Soumendra Padhi',
'Srijit Mukherji',
'Tharun Bhascker Dhaassyam'

```
df_directors=df_india_movies.groupby(['Directors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
df_directors=df_directors[df_directors['Directors']!='Unknown Director']
plt.figure(figsize=(15,8))
plt.barh(df_directors[::-1]['Directors'], df_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Directors')
plt.show()
```

Farah Khan	4
Farah Nabulsi	1
Farah X	1
Farhad Safinia	1
Farhan Akhtar	4
Farhan Alam	1
Faris Godus	2
Farishad I. Latjuba	1
Farjad Nabi	1
Fazila Allana	1
Federico D'Alessandro	1
Felix Herngren	1
Felix Starck	1
Felix Thompson	1
Felix Van Groeningen	1
Felix van Groeningen	1
Fellipe Barbosa	1
Femi D. Ogunsanwo	1
Femi Oyeniran	2
Feng Xiaogang	1
Ferdinando Cito Filomarino	1
Fermín Muguruza	1
Fernando Ayllón	8

Alison Mandel	1
Alison Pill	8
Alison Retzloff	1
Alison Thornton	1
Alison Wandzura	1
Alissa Latow	1
Alisson Santiago	1
Alistair Abell	2
Alistair McGowan	2
Alistair Petrie	2
Aliton Graça	1
Alix Macey	1
Aljin Abella	1
Alka Amin	2
a Badola Kaushal	4
Alka Chawla	1
Alki David	1
Alkoya Brunson	3
Alla Tumanian	1
Allan Corduner	1
Allan Graf	1
Allan Hawco	2
Allan Hyde	1
Allan Kournikova	1
Allan Lima	1



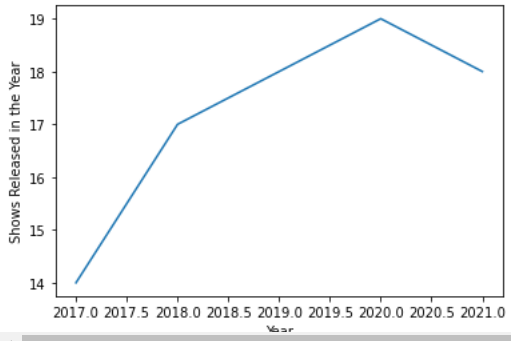

```
df_directors['Directors'].values
```

```
array(['Rajiv Chilaka', 'Suhas Kadav', 'David Dhawan', 'Anurag Kashyap',
       'Umesh Mehra', 'Ram Gopal Varma', 'Dibakar Banerjee',
       'Zoya Akhtar', 'Tilak Shetty', 'Rajkumar Santoshi', 'Priyadarshan',
       'Sooraj R. Barjatya', 'Ashutosh Gowariker', 'Milan Luthria'],
      dtype=object)
```

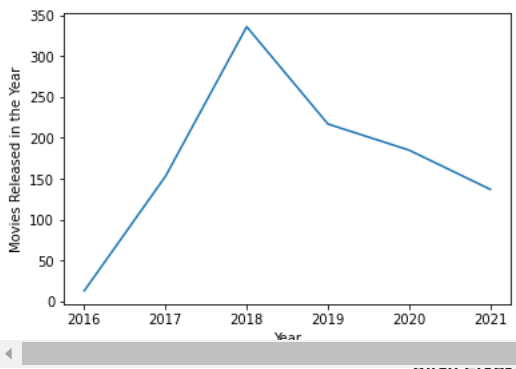
Popular directors across movies in India:-

'Rajiv Chilaka',
'Suhas Kadav',
'David Dhawan',
'Umesh Mehra',
'Anurag Kashyap',
'Ram Gopal Varma',
'Dibakar Banerjee',
'Zoya Akhtar',
'Tilak Shetty',
'Rajkumar Santoshi',
'Priyadarshan',
'Sooraj R. Barjatya',
'Ashutosh Gowariker',
'Milan Luthria'

```
df_year=df_india_shows.groupby(['year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_year, x='year', y='title')
plt.ylabel("Shows Released in the Year")
plt.xlabel("Year")
plt.show()
```



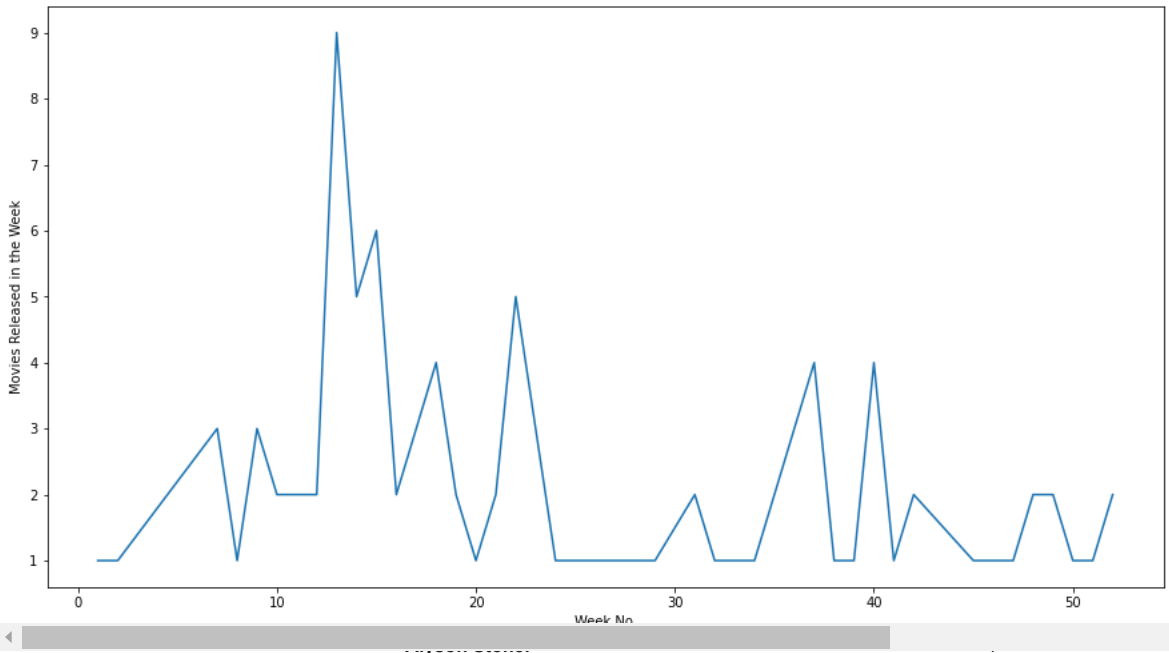

```
df_year=df_india_movies.groupby(['year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_year, x='year', y='title')
plt.ylabel("Movies Released in the Year")
plt.xlabel("Year")
plt.show()
```



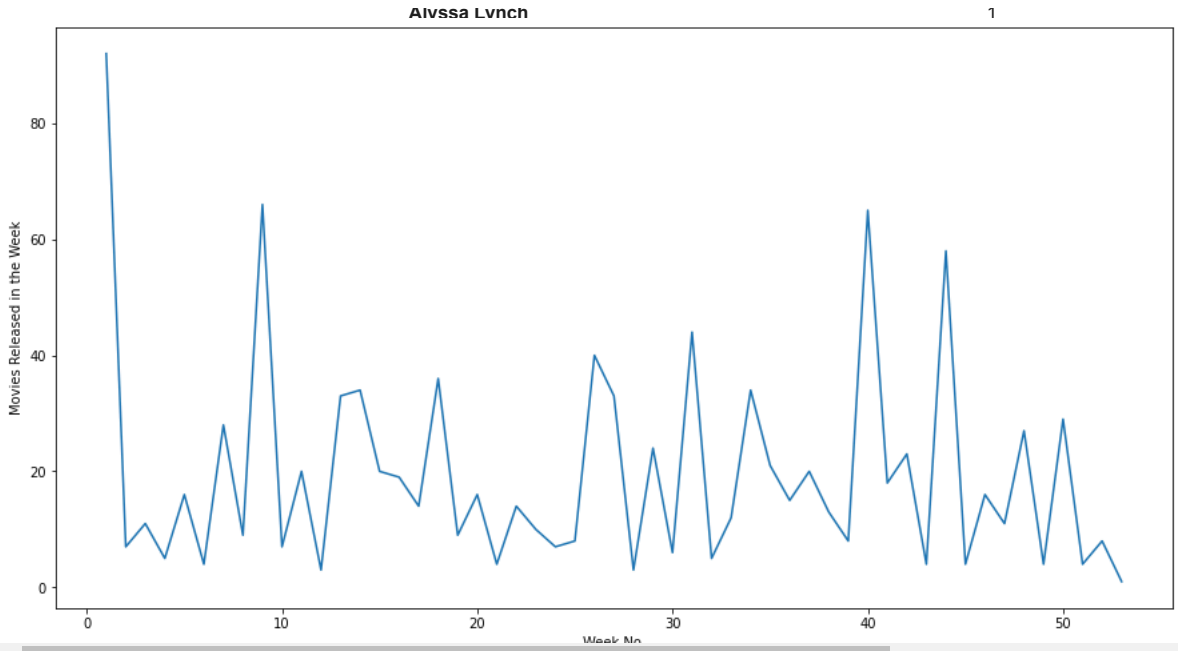
In India,TV Shows were increasingly being added till 2020, though the addition of shows reduced in 2021.

In India, Movies were increasingly added till 2018 but it has been a huge downhill since then. Now that's preposterous, since and soemthing has to be recommended to the Netflix Team with regards to that.

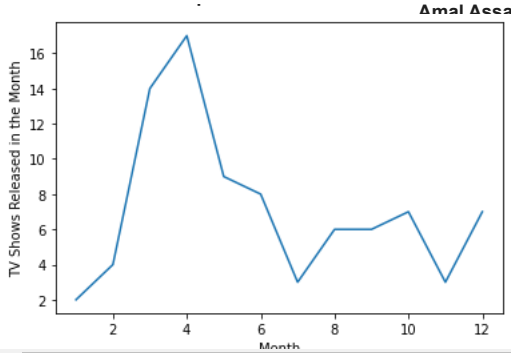
```
df_week=df_india_shows.groupby(['week_Added']).agg({"title":"nunique"}).reset_index()
plt.figure(figsize=(15,8))
sns.lineplot(data=df_week, x='week_Added', y='title')
plt.ylabel("Movies Released in the Week")
plt.xlabel("Week No.")
plt.show()
```



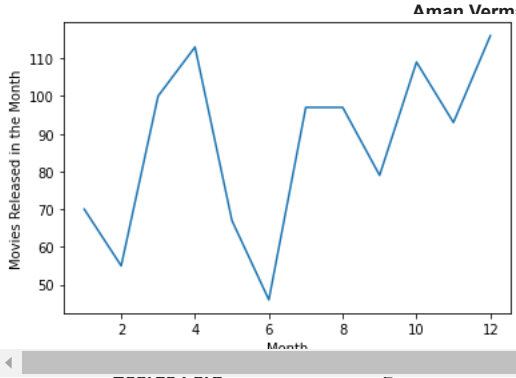
```
df_week=df_india_movies.groupby(['week_Added']).agg({"title":"nunique"}).reset_index()
plt.figure(figsize=(15,8))
sns.lineplot(data=df_week, x='week_Added', y='title')
plt.ylabel("Movies Released in the Week")
plt.xlabel("Week No.")
plt.show()
```




```
df_month=df_india_shows.groupby(['month_added']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_month, x='month_added', y='title')
plt.ylabel("TV Shows Released in the Month")
plt.xlabel("Month")
plt.show()
```



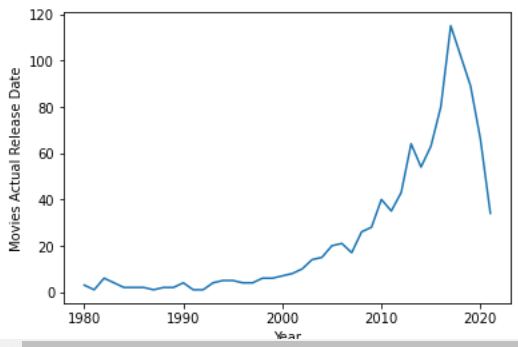

```
df_month=df_india_movies.groupby(['month_added']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_month, x='month_added', y='title')
plt.ylabel("Movies Released in the Month")
plt.xlabel("Month")
plt.show()
```



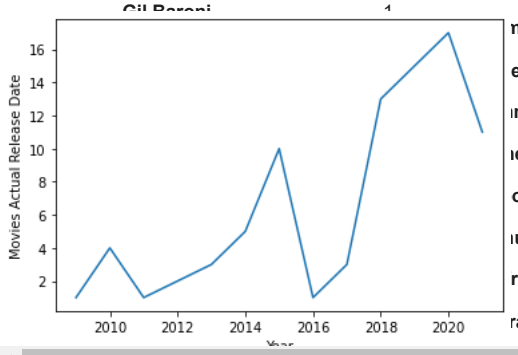
TV Shows are added in Netflix by a tremendous amount in April in India

Movies are added in Netflix in India by a tremendous amount in first week/last month of current year and first month of next year

```
df_release_year=df_india_movies[df_india_movies['release_year']>=1980].groupby(['release_year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_release_year, x='release_year', y='title')
plt.ylabel("Movies Actual Release Date")
plt.xlabel("Year")
plt.show()
```




```
df_release_year=df_india_shows[df_india_shows['release_year']>=1980].groupby(['release_year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_release_year, x='release_year', y='title')
plt.ylabel("Movies Actual Release Date")
plt.xlabel("Year")
plt.show()
```



The understandable trend amongs movies and TV Shows across India in Netflix is the reduction of movies after 2020

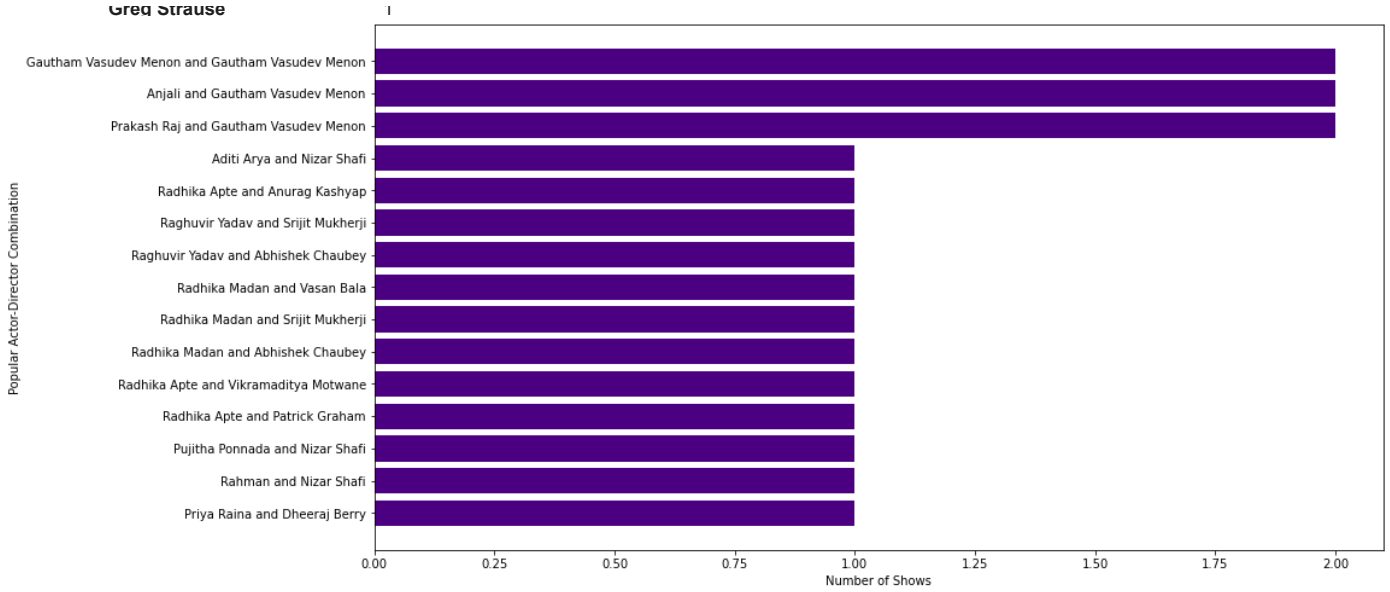
```
#Analysing a combination of actors and directors
df_india_movies['Actor_Director_Combination'] = df_india_movies.Actors.str.cat(df_india_movies.Directors, sep=' and ')
df_india_movies_subset=df_india_movies[df_india_movies['Actors']!='Unknown Actor']
df_india_movies_subset=df_india_movies_subset[df_india_movies_subset['Directors']!='Unknown Director']
df_india_movies_subset.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	Modified_Ac
621	Avvai Shanmughi	Kamal Hassan	K.S. Ravikumar	Comedies	India	s23	Movie	September 21, 2021	1996	TV-PG	150-200	2
622	Avvai Shanmughi	Kamal Hassan	K.S. Ravikumar	International Movies	India	s23	Movie	September 21, 2021	1996	TV-PG	150-200	2
629	Avvai Shanmughi	Nassar	K.S. Ravikumar	Comedies	India	s23	Movie	September 21, 2021	1996	TV-PG	150-200	2
630	Avvai Shanmughi	Nassar	K.S. Ravikumar	International Movies	India	s23	Movie	September 21, 2021	1996	TV-PG	150-200	2
631	Avvai Shanmughi	S.P. Balasubrahmanyam	K.S. Ravikumar	Comedies	India	s23	Movie	September 21, 2021	1996	TV-PG	150-200	2

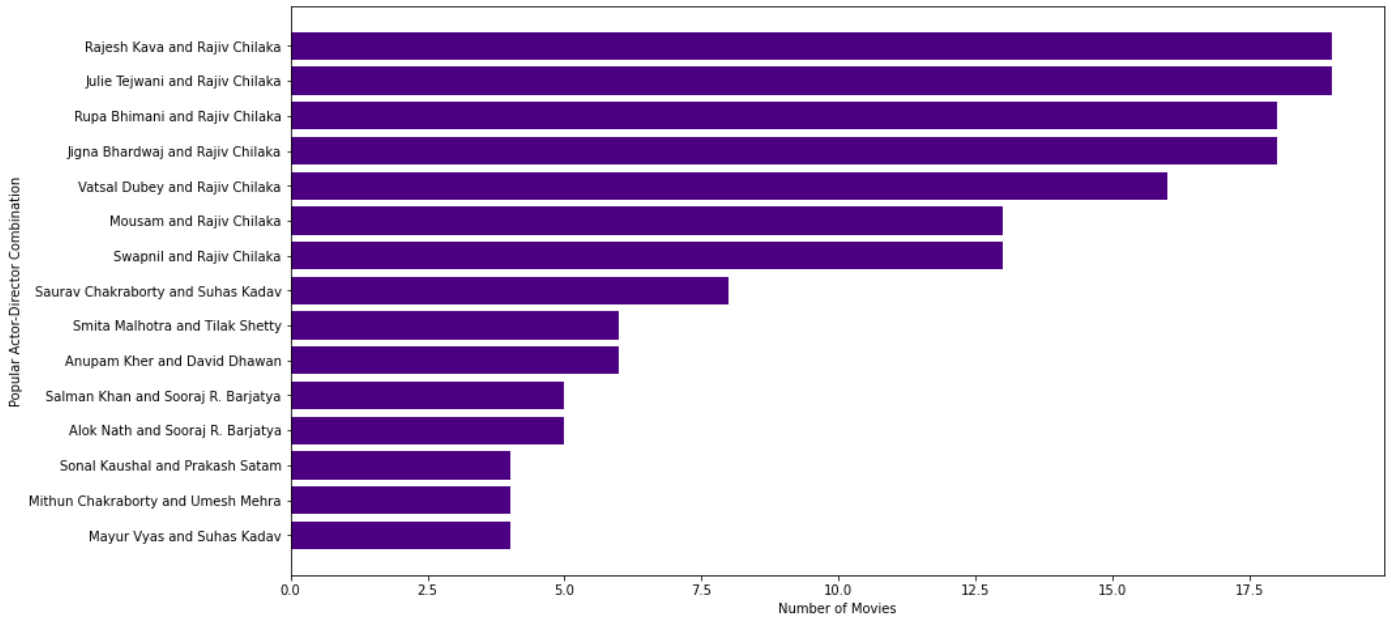
```
df_india_shows['Actor_Director_Combination'] = df_india_shows.Actors.str.cat(df_india_shows.Directors, sep=' and ')
df_india_shows_subset=df_india_shows[df_india_shows['Actors']!='Unknown Actor']
df_india_shows_subset=df_india_shows_subset[df_india_shows_subset['Directors']!='Unknown Director']
df_india_shows_subset.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	Modified_Added_Date	mont
7005	Navarasa	Suriya	Bejoy Nambiar	TV Shows	India	s298	TV Show	August 6, 2021	2021	TV-MA	1 Season	2021-08-06	
7006	Navarasa	Suriya	Priyadarshan	TV Shows	India	s298	TV Show	August 6, 2021	2021	TV-MA	1 Season	2021-08-06	
7007	Navarasa	Suriya	Karthik Narain	TV Shows	India	s298	TV Show	August 6, 2021	2021	TV-MA	1 Season	2021-08-06	
7008	Navarasa	Suriya	Vasanth Sai	TV Shows	India	s298	TV Show	August 6, 2021	2021	TV-MA	1 Season	2021-08-06	
7009	Navarasa	Suriya	Karthik Subbaraj	TV Shows	India	s298	TV Show	August 6, 2021	2021	TV-MA	1 Season	2021-08-06	

```
df_actors_directors=df_india_shows_subset.groupby(['Actor_Director_Combination']).agg({"title":"nunique"}).reset_index().sort_values(by=['ti
plt.figure(figsize=(15,8))
plt.barh(df_actors_directors[::-1]['Actor_Director_Combination'], df_actors_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Shows')
plt.ylabel('Popular Actor-Director Combination')
plt.show()
```




```
df_actors_directors=df_india_movies_subset.groupby(['Actor_Director_Combination']).agg({"title":"nunique"}).reset_index().sort_values(by=['t
plt.figure(figsize=(15,8))
plt.barh(df_actors_directors[::-1]['Actor_Director_Combination'], df_actors_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Actor-Director Combination')
plt.show()
```




```
df_india_movies[df_india_movies['Directors']=='Rajiv Chilaka']
```

Hamisha Daryani Ahuja	1
Han Han	1
Han Kwang Il	1
Han Qing	1
Han Yan	1
Handan İpekçi	1
Hang-Jun Jang	1
Hani Al-Shaibani	1
Hani Hamdi	2
Hannah Fidell	2
Hannes Stöhr	1
Hanno Olderdissen	1
Hanns-Bruno Kammertöns	1
Hans Petter Moland	1
Hanung Bramantyo	8
Hany Motawie	1
Hao Wu	1
Haoling Li	1
Harald Zwart	1
Hardik Mehta	3
Hari Nath	1
Hari Viswanath	1
Harjit Singh	1
Hark Tsui	1
Harmony Korine	1
Harold Becker	1
Harold Cronk	2
Harpeet Singh	1
Harrison Smith	1

Amir Bashir	1
Amir El Kacem	1
Amir El-Masry	1
Amir O'Neil	1
Amir Raza Hussain	1
Amir Salah	1
Amir Talai	1
Amir Tessler	1
Amir Wilson	1
Amira Casar	2
Amirah Vann	2
Amit Behl	2
Amit Deondi	2
Amit Kumar	1
Amit Kumar Tiwari	1
Amit Pachori	1
Amit Rahav	1
Amit Sadh	2
Amit Shah	2
Amit Sial	6
Amit Tandon	2
Amit Varma	1
Amita Khopkar	1
Amita Pathak	1
Amita Suman	2
nitabh Bachchan	28
Amitabh Srivastava	1
abha Bhattacharjee	1
Amith Chakalakkal	1

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	Modified_Added_
10058	Chhota Bheem - Neeli Pahaadi	Vatsal Dubey	Rajiv Chilaka	Children & Family Movies	India	s407	Movie	July 22, 2021	2013	TV-Y7	50-80	2021-
10059	Chhota Bheem - Neeli Pahaadi	Julie Tejwani	Rajiv Chilaka	Children & Family Movies	India	s407	Movie	July 22, 2021	2013	TV-Y7	50-80	2021-
10060	Chhota Bheem - Neeli Pahaadi	Rupa Bhimani	Rajiv Chilaka	Children & Family Movies	India	s407	Movie	July 22, 2021	2013	TV-Y7	50-80	2021-
10061	Chhota Bheem - Neeli Pahaadi	Jigna Bhardwaj	Rajiv Chilaka	Children & Family Movies	India	s407	Movie	July 22, 2021	2013	TV-Y7	50-80	2021-
10062	Chhota Bheem - Neeli Pahaadi	Rajesh Kava	Rajiv Chilaka	Children & Family Movies	India	s407	Movie	July 22, 2021	2013	TV-Y7	50-80	2021-
10063	Chhota Bheem - Neeli Pahaadi	Mousam	Rajiv Chilaka	Children & Family Movies	India	s407	Movie	July 22, 2021	2013	TV-Y7	50-80	2021-
10064	Chhota Bheem - Neeli Pahaadi	Swapnil	Rajiv Chilaka	Children & Family Movies	India	s407	Movie	July 22, 2021	2013	TV-Y7	50-80	2021-
10065	Chhota Bheem & Ganesh	Vatsal Dubey	Rajiv Chilaka	Children & Family Movies	India	s408	Movie	July 22, 2021	2009	TV-Y7	50-80	2021-
10066	Chhota Bheem & Ganesh	Julie Tejwani	Rajiv Chilaka	Children & Family Movies	India	s408	Movie	July 22, 2021	2009	TV-Y7	50-80	2021-
10067	Chhota Bheem & Ganesh	Rupa Bhimani	Rajiv Chilaka	Children & Family Movies	India	s408	Movie	July 22, 2021	2009	TV-Y7	50-80	2021-
10068	Chhota Bheem & Ganesh	Jigna Bhardwaj	Rajiv Chilaka	Children & Family Movies	India	s408	Movie	July 22, 2021	2009	TV-Y7	50-80	2021-
10069	Chhota Bheem & Ganesh	Rajesh Kava	Rajiv Chilaka	Children & Family Movies	India	s408	Movie	July 22, 2021	2009	TV-Y7	50-80	2021-
10071	Chhota Bheem & Ganesh	Swapnil	Rajiv Chilaka	Children & Family Movies	India	s408	Movie	July 22, 2021	2009	TV-Y7	50-80	2021-
10072	Chhota Bheem & Krishna: Mayanagari	Vatsal Dubey	Rajiv Chilaka	Children & Family Movies	India	s409	Movie	July 22, 2021	2011	TV-Y7	50-80	2021-
10073	Chhota Bheem & Krishna: Mayanagari	Julie Tejwani	Rajiv Chilaka	Children & Family Movies	India	s409	Movie	July 22, 2021	2011	TV-Y7	50-80	2021-
10074	Chhota Bheem & Krishna: Mayanagari	Rupa Bhimani	Rajiv Chilaka	Children & Family Movies	India	s409	Movie	July 22, 2021	2011	TV-Y7	50-80	2021-

It seems that Rajiv Chilaka has worked on Chota Bheem and has been able to create some good content in its movies. He can be relied on for more Chota Bheem stories

```
df_actors_directors['Actor_Director_Combination'].values
```

```
array(['Rajesh Kava and Rajiv Chilaka', 'Julie Tejwani and Rajiv Chilaka',
       'Rupa Bhimani and Rajiv Chilaka', 'Jigna Bhardwaj and Rajiv Chilaka',
       'Vatsal Dubey and Rajiv Chilaka', 'Mousam and Rajiv Chilaka',
       'Swapnil and Rajiv Chilaka', 'Saurav Chakraborty and Suhas Kadav',
       'Smita Malhotra and Tilak Shetty', 'Jaideep Kher and David Dhawan',
       'Salman Khan and Sooraj R. Barjatya',
       'Alok Nath and Sooraj R. Barjatya',
       'Smaal Kaushal and Prakash Satam',
       'Mithun Chakraborty and Umesh Mehra', 'Mayur Vyas and Suhas Kadav'],
      dtype=object)
```

The Most Popular Actor Director Combination in Movies across India are:-

'Rajesh Kava and Rajiv Chilaka',

'Julie Tejwani and Rajiv Chilaka',

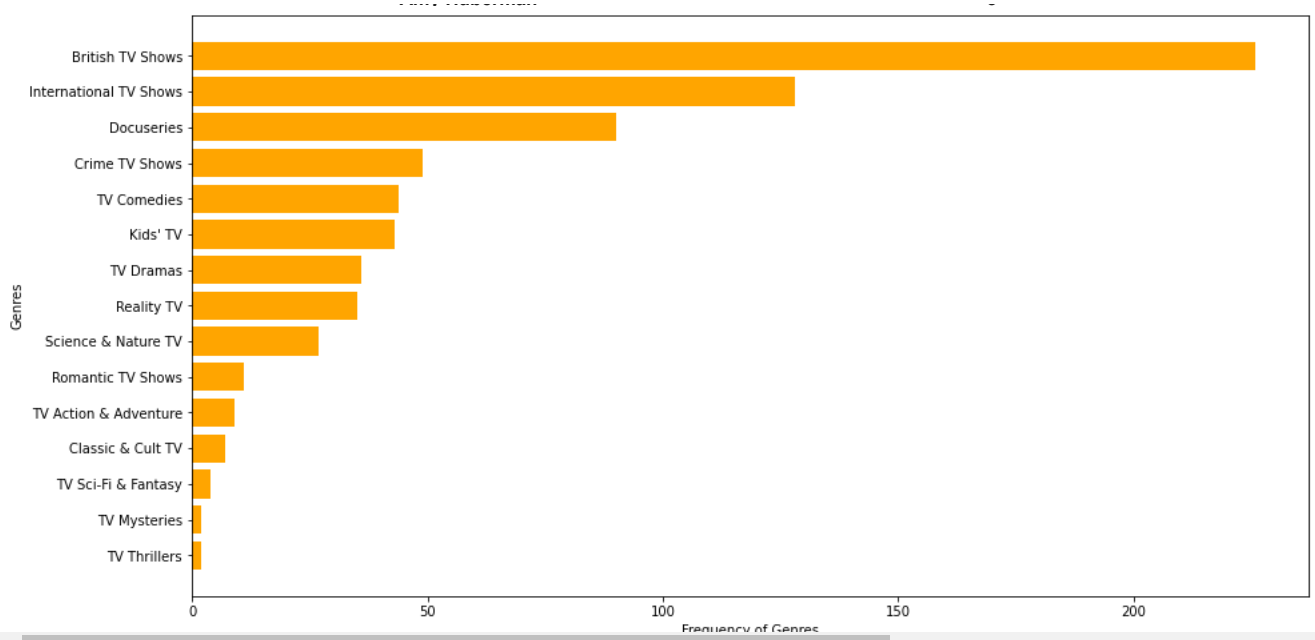
'Rupa Bhimani and Rajiv Chilaka',

'Jigna Bhardwaj and Rajiv Chilaka',
'Vatsal Dubey and Rajiv Chilaka',
'Mousam and Rajiv Chilaka',
'Swapnil and Rajiv Chilaka',
'Saurav Chakraborty and Suhas Kadav',
'Smita Malhotra and Vrak Shetty',
'Anupam Kher and Rajiv Chilaka',
'Salman Khan and Prabhu Deva'

Univariate Analysis separately for shows and movies in United Kingdom

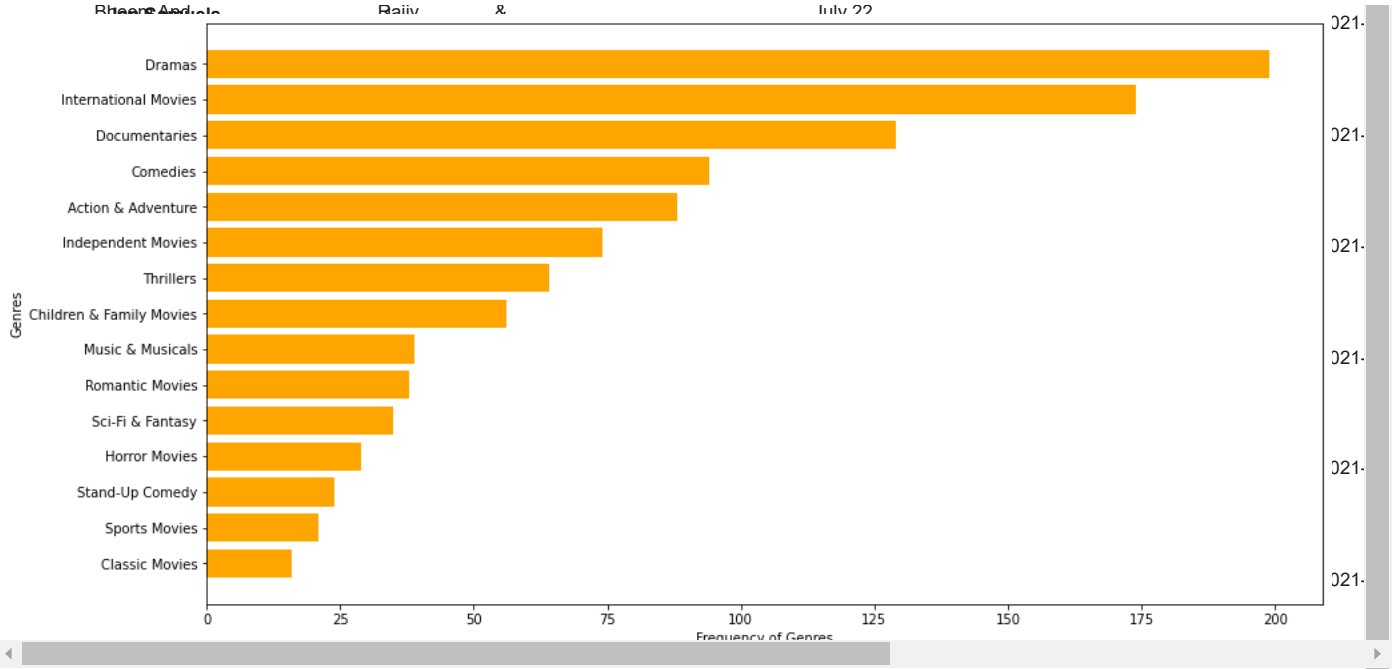
```
#Analyzing India for both shows and movies
df_uk_shows=df_final1[df_final1['country']=='United Kingdom'][df_final1[df_final1['country']=='United Kingdom']['type']=='TV Show']
df_uk_movies=df_final1[df_final1['country']=='United Kingdom'][df_final1[df_final1['country']=='United Kingdom']['type']=='Movie']
```

```
df_genre=df_uk_shows.groupby(['Genre']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_genre[::-1]['Genre'], df_genre[::-1]['title'],color=['orange'])
plt.xlabel('Frequency of Genres')
plt.ylabel('Genres')
plt.show()
```



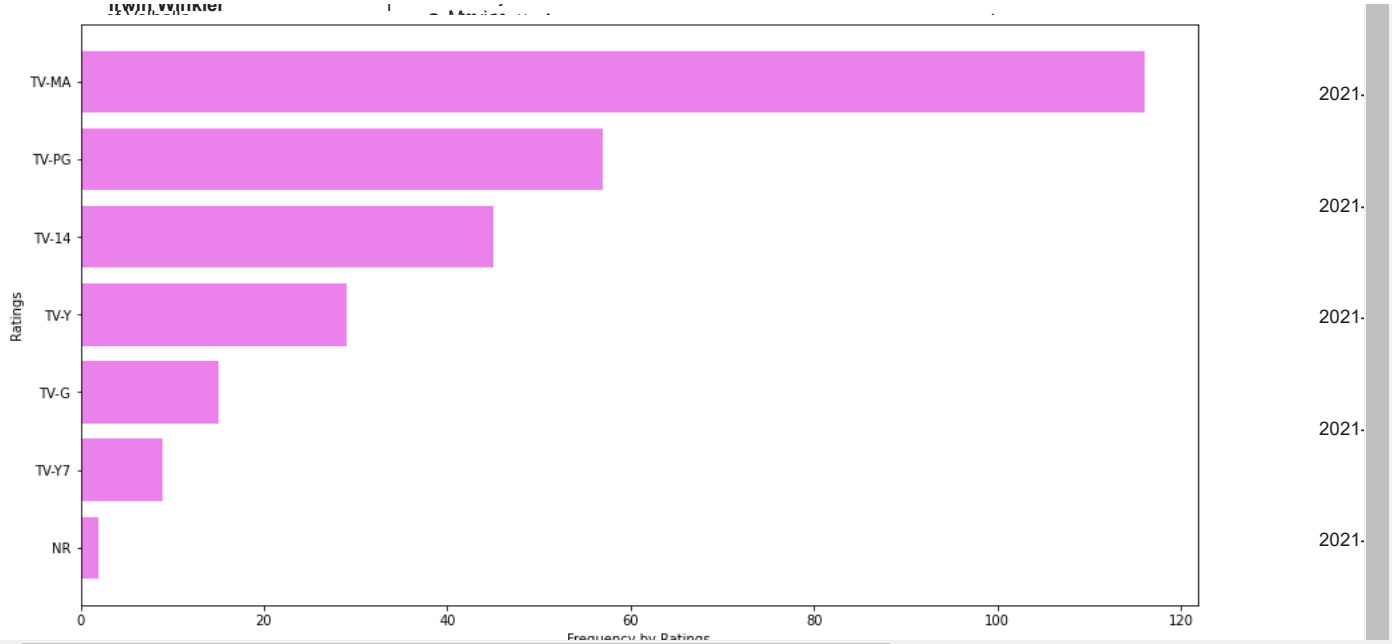
British TV Shows,International TV Shows,Docuseries, Crime, Comedy are widely watched Genres in TV Shows in UK

```
df_genre=df_uk_movies.groupby(['Genre']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_genre[::-1]['Genre'], df_genre[::-1]['title'],color=['orange'])
plt.xlabel('Frequency of Genres')
plt.ylabel('Genres')
plt.show()
```

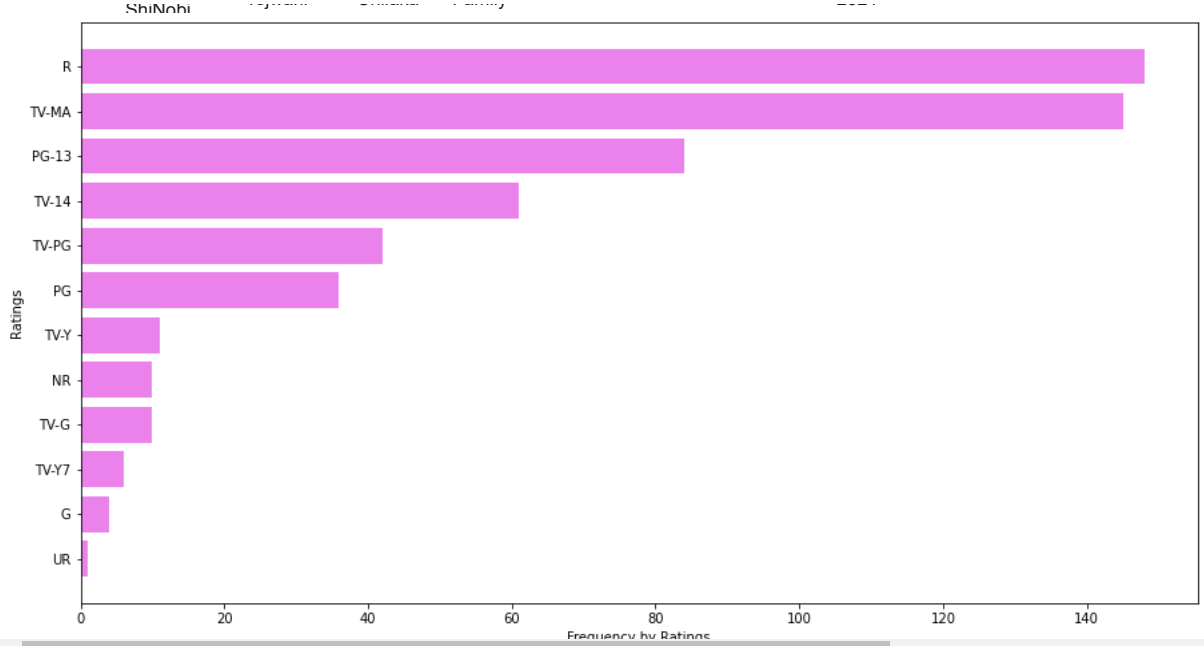


International Movies,Drama,Comedy,Indpeendent Movies and Action, Romance Genres in Movies are prevalent in UK

```
df_rating=df_uk_shows.groupby(['rating']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_rating[::-1]['rating'], df_rating[::-1]['title'],color=['violet'])
plt.xlabel('Frequency by Ratings')
plt.ylabel('Ratings')
plt.show()
```

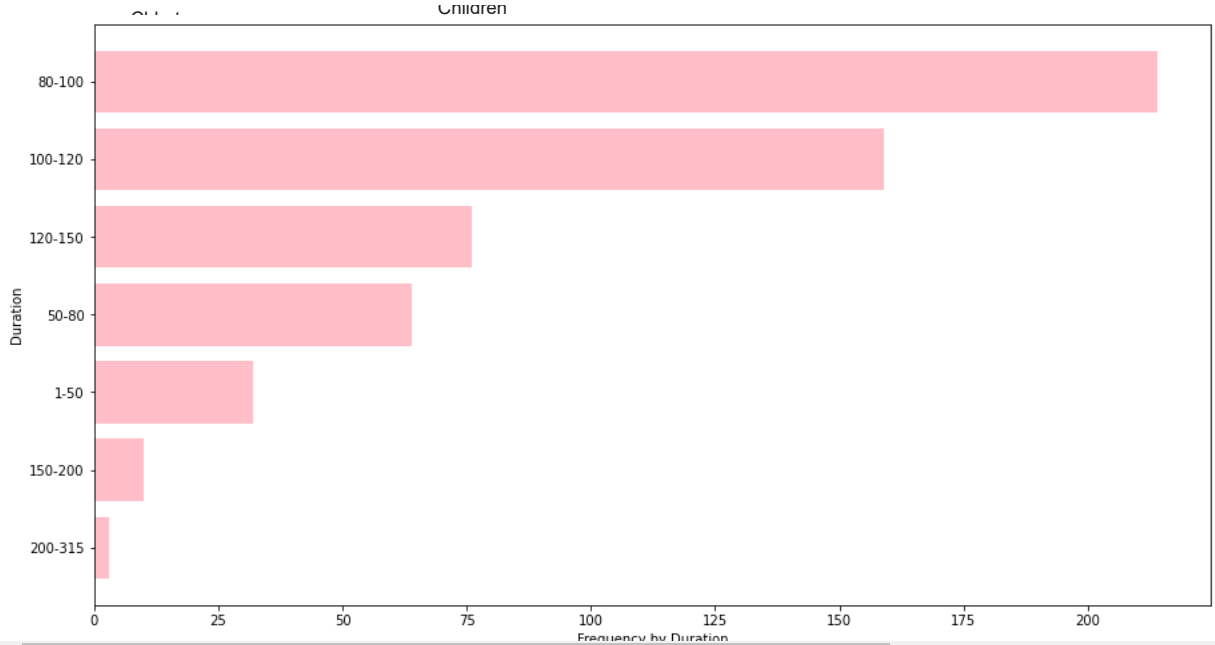



```
df_rating=df_uk_movies.groupby(['rating']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
plt.figure(figsize=(15,8))
plt.barh(df_rating[::-1]['rating'], df_rating[::-1]['title'],color=['violet'])
plt.xlabel('Frequency by Ratings')
plt.ylabel('Ratings')
plt.show()
```



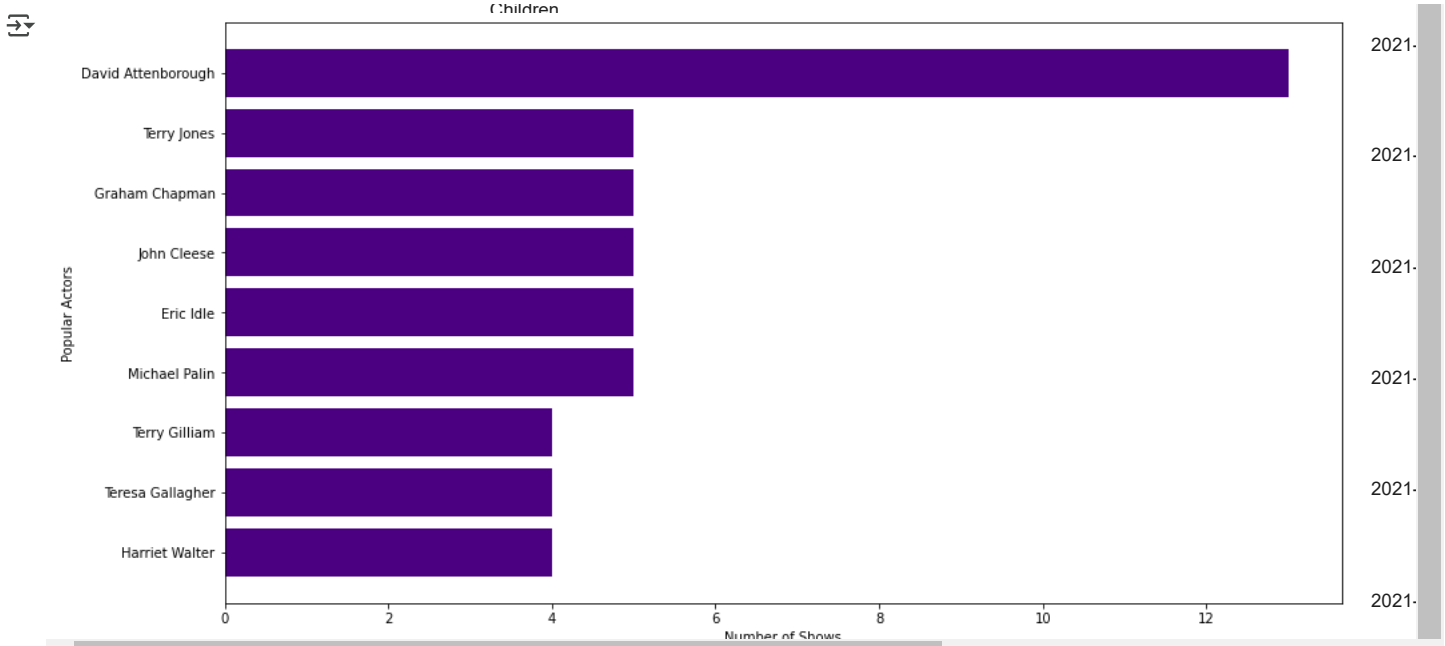
So it seems plausible to conclude that the popular ratings across Netflix includes Parental Guidance and Mature Audiences in TV Shows and R Rated+MA Rated in Movies in UK

```
df_duration=df_uk_movies.groupby(['duration']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
plt.figure(figsize=(15,8))
plt.barh(df_duration[::-1]['duration'], df_duration[::-1]['title'],color=['pink'])
plt.xlabel('Frequency by Duration')
plt.ylabel('Duration')
plt.show()
```



Across Movies ranges of minutes in UK have a sweet spot at 80-120 mins.

```
df_actors=df_uk_shows.groupby(['Actors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
df_actors=df_actors[df_actors['Actors']!='Unknown Actor']
plt.figure(figsize=(15,8))
plt.barh(df_actors[::-1]['Actors'], df_actors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Shows')
plt.ylabel('Popular Actors')
plt.show()
```



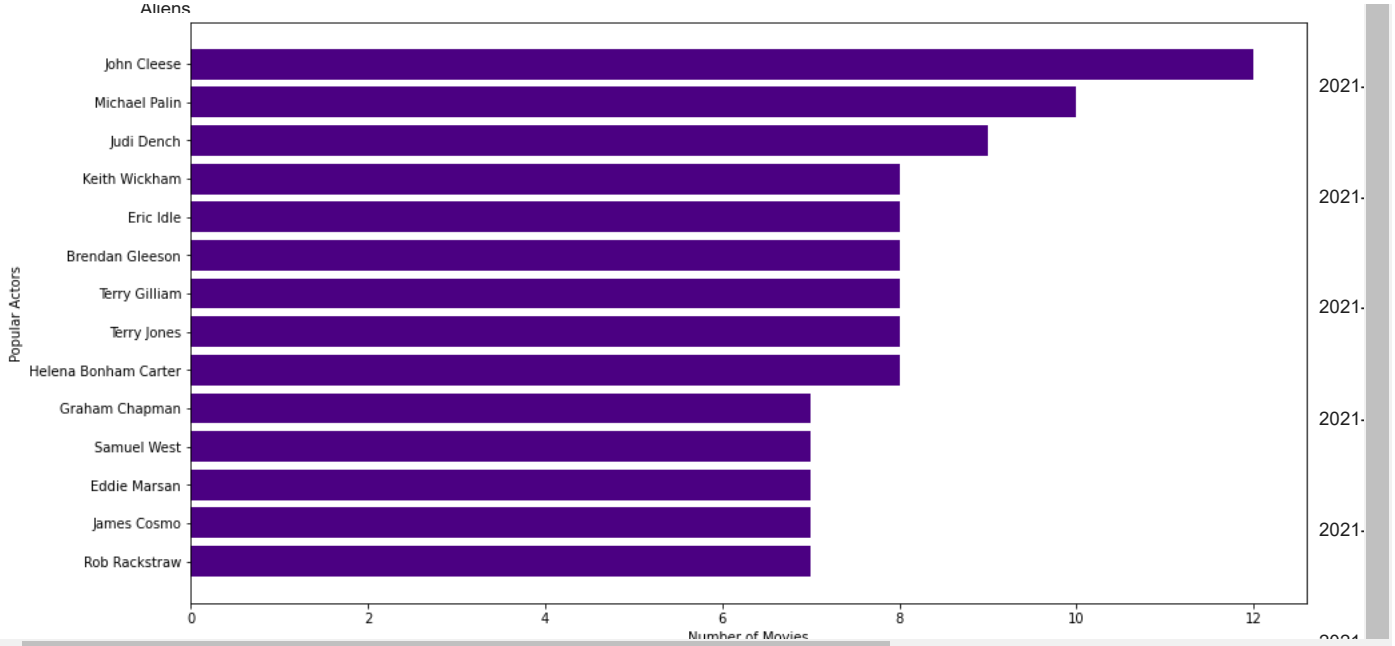

```
df_actors['Actors'].values
```

```
array(['David Attenborough', 'Terry Jones', 'Graham Chapman',
       'John Cleese', 'Eric Idle', 'Michael Palin', 'Terry Gilliam',
       'Teresa Gallagher', 'Harriet Walter'], dtype=object)
```

Popular Actors in TV Shows in UK are:-

'David Attenborough',
'Terry Jones',
'Graham Chapman',
'John Cleese',
'Eric Idle',
'Michael Palin',
'Terry Gilliam',
'Teresa Gallagher',
'Harriet Walter'

```
df_actors=df_uk_movies.groupby(['Actors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
df_actors=df_actors[df_actors['Actors']!='Unknown Actor']
plt.figure(figsize=(15,8))
plt.barh(df_actors[::-1]['Actors'], df_actors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Actors')
plt.show()
```



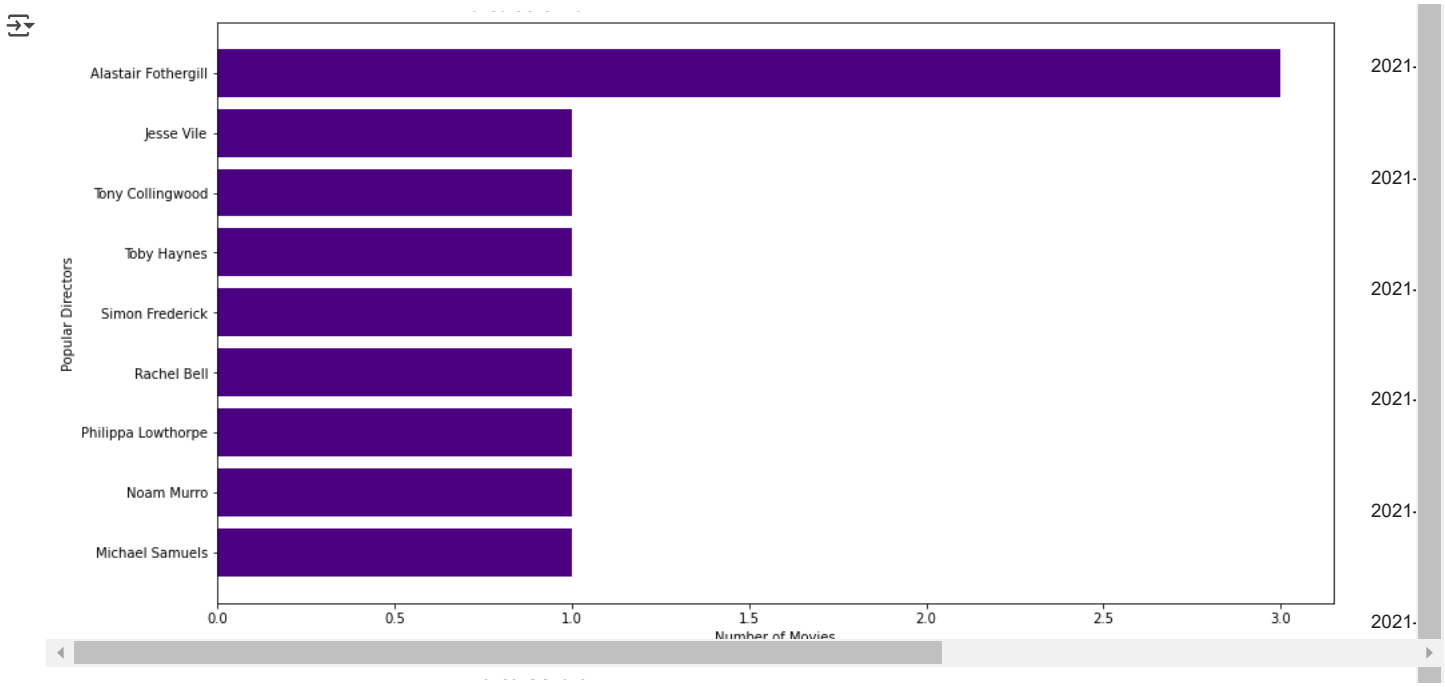

```
df_actors['Actors'].values
```

```
array(['John Cleese', 'Michael Palin', 'Judi Dench', 'Keith Wickham',
       'Eric Idle', 'Brendan Gleeson', 'Terry Gilliam', 'Terry Jones',
       'Helena Bonham Carter', 'Graham Chapman', 'Samuel West',
       'Eddie Marsan', 'James Cosmo', 'Rob Rackstraw'], dtype=object)
```

Popular actors across Movies in UK:

'John Cleese',
'Michael Palin',
'Judi Dench',
'Keith Wickham',
'Eric Idle',
'Brendan Gleeson',
'Terry Gilliam',
'Terry Jones',
'Helena Bonham Carter',
'Graham Chapman',
'Samuel West',
'Eddie Marsan',
'James Cosmo',
'Rob Rackstraw'

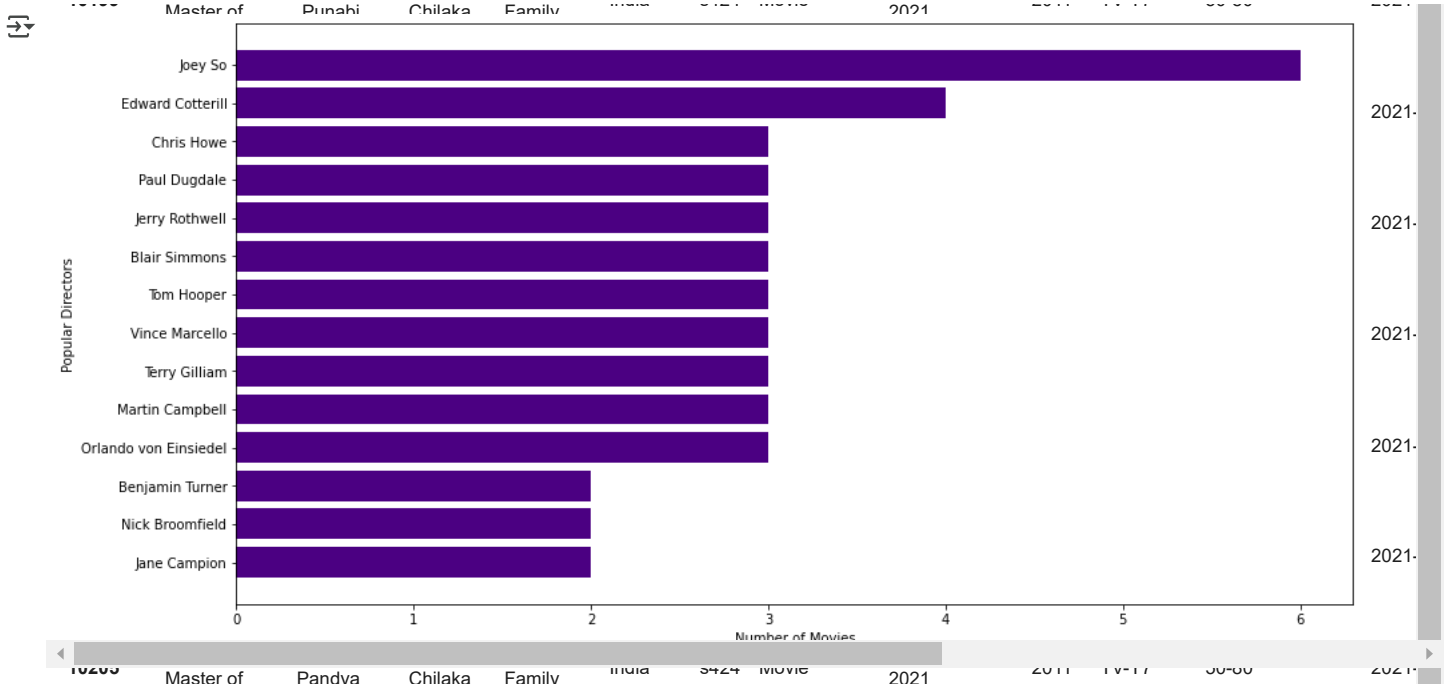
```
df_directors=df_uk_shows.groupby(['Directors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:10]
df_directors=df_directors[df_directors['Directors']!='Unknown Director']
plt.figure(figsize=(15,8))
plt.barh(df_directors[::-1]['Directors'], df_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Directors')
plt.show()
```




```
df_directors['Directors'].values
```

```
array(['Alastair Fothergill', 'Jesse Vile', 'Tony Collingwood',
       'Toby Haynes', 'Simon Frederick', 'Rachel Bell',
       'Philippa Lowthorpe', 'Noam Murro', 'Michael Samuels'],
      dtype=object)
```

```
df_directors=df_uk_movies.groupby(['Directors']).agg({"title":"nunique"}).reset_index().sort_values(by=['title'],ascending=False)[:15]
df_directors=df_directors[df_directors['Directors']!='Unknown Director']
plt.figure(figsize=(15,8))
plt.barh(df_directors[::-1]['Directors'], df_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Directors')
plt.show()
```



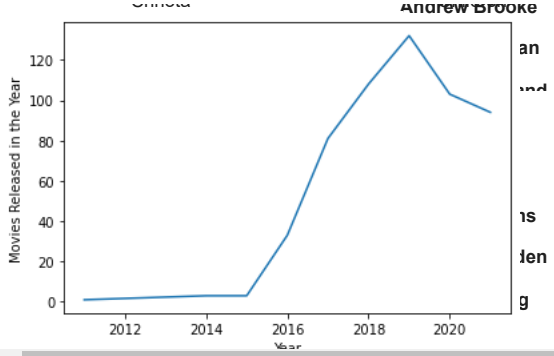

```
df_directors['Directors'].values
```

```
array(['Joey So', 'Edward Cotterill', 'Chris Howe', 'Paul Dugdale',
       'Jerry Rothwell', 'Blair Simmons', 'Tom Hooper', 'Vince Marcello',
       'Terry Gilliam', 'Martin Campbell', 'Orlando von Einsiedel',
       'Benjamin Turner', 'Nick Broomfield', 'Jane Campion'], dtype=object)
```

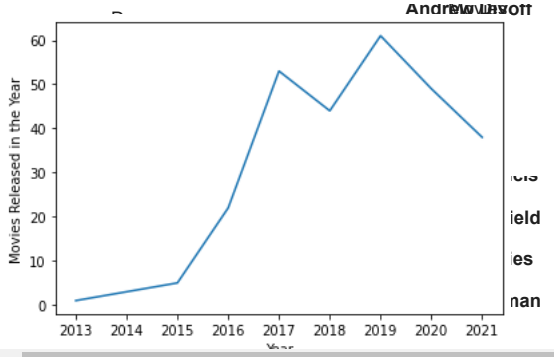
Popular directors across movies in UK:-

'Joey So',
'Edward Gotterill'

```
df_year=df_uk_movies.groupby(['year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_year, x='year', y='title')
plt.ylabel("Movies Released in the Year")
plt.xlabel("Year")
plt.show()
```



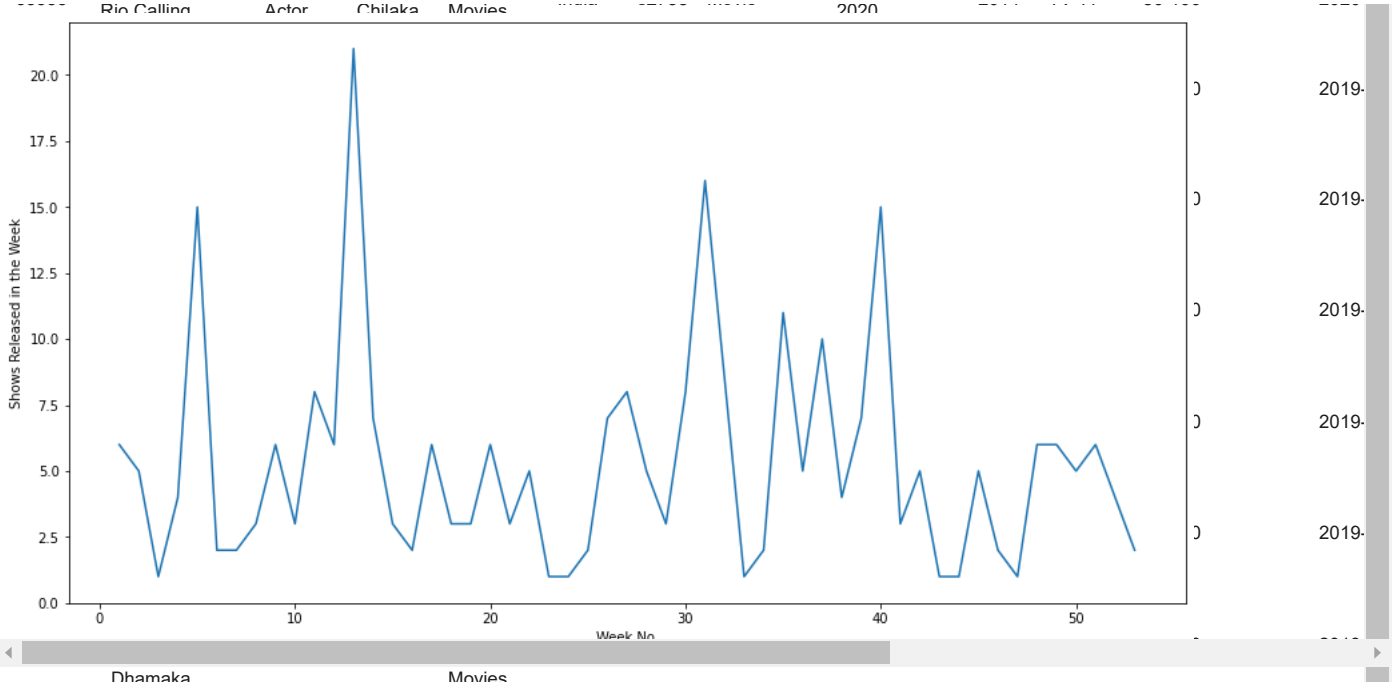

```
df_year=df_uk_shows.groupby(['year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_year, x='year', y='title')
plt.ylabel("Movies Released in the Year")
plt.xlabel("Year")
plt.show()
```



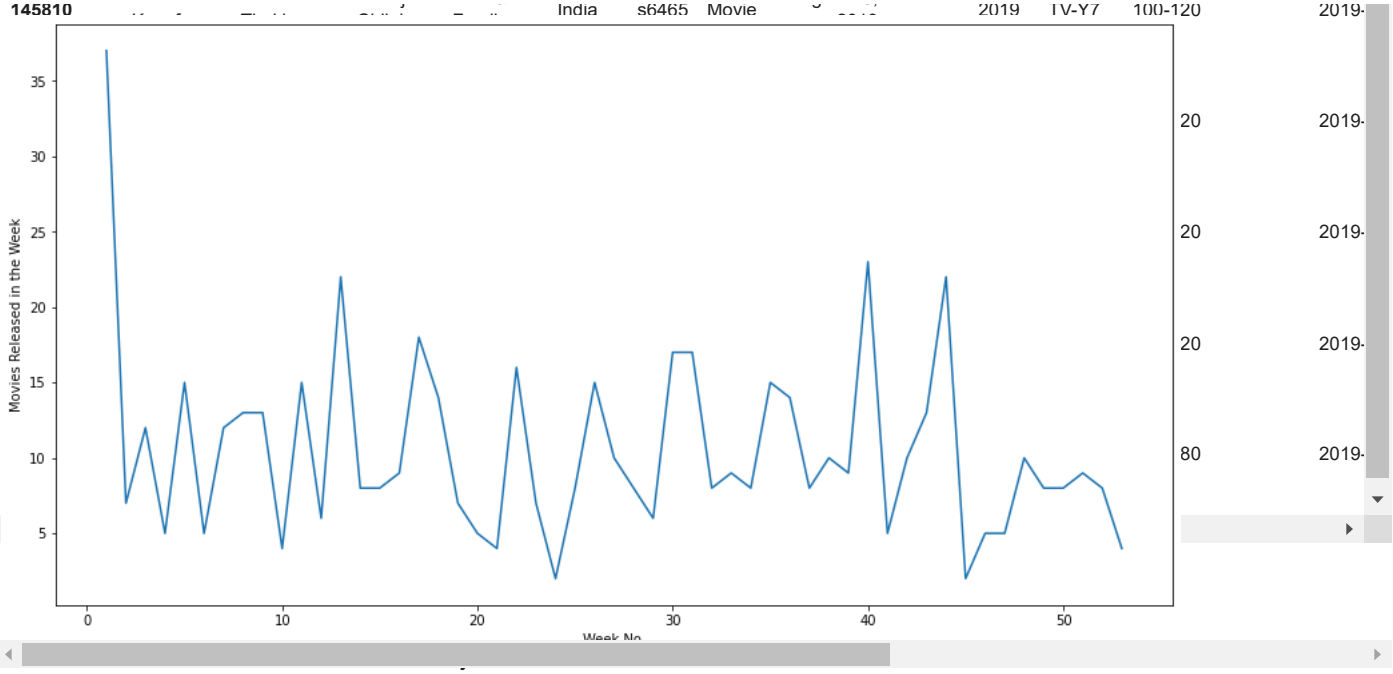
In terms of TV Shows, UK saw a downfall in 2018 from 2017, then a great increase in 2019 but has been reducing since then.

In terms of Movies, the number of popular movies in UK increased till 2019, since then it's decreasing.

```
df_week=df_uk_shows.groupby(['week_Added']).agg({"title":"nunique"}).reset_index()
plt.figure(figsize=(15,8))
sns.lineplot(data=df_week, x='week_Added', y='title')
plt.ylabel("Shows Released in the Week")
plt.xlabel("Week No.")
plt.show()
```

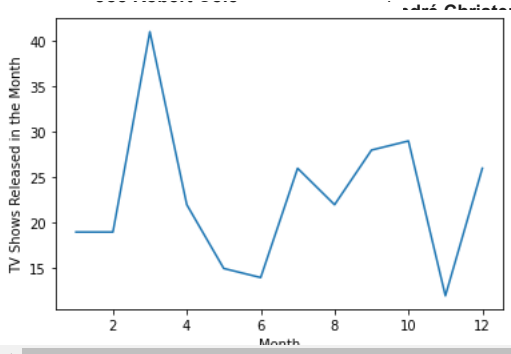



```
df_week=df_uk_movies.groupby(['week_Added']).agg({"title":"nunique"}).reset_index()
plt.figure(figsize=(15,8))
sns.lineplot(data=df_week, x='week_Added', y='title')
plt.ylabel("Movies Released in the Week")
plt.xlabel("Week No.")
plt.show()
```

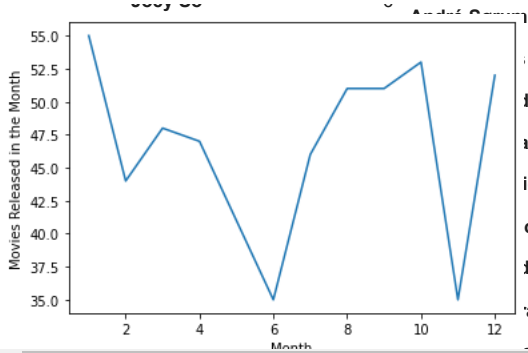



```
df_month=df_uk_shows.groupby(['month_added']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_month, x='month_added', y='title')
plt.ylabel("TV Shows Released in the Month")
plt.xlabel("Month")
plt.show()
```

Joe Menendez	1	**Andrzej Mrowiec**	1
Joe Miale	1	**Andrzej Seweryn**	2
Joe Murray	1	**Andrzej Łapicki**	1
Joe Nussbaum	1	**András Faragó**	1
Joe Penna	1	**André Bharti**	1
Joe Piscatella	1	**André Chiang**	1
Joe Robert Cole	1		



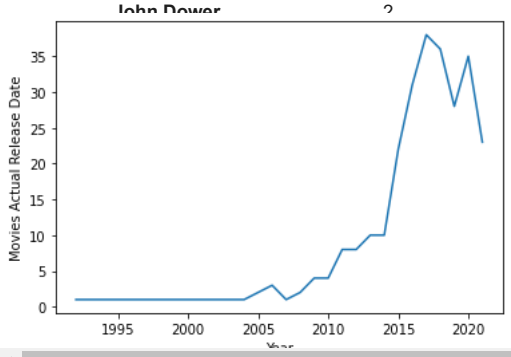

```
df_month=df_uk_movies.groupby(['month_added']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_month, x='month_added', y='title')
plt.ylabel("Movies Released in the Month")
plt.xlabel("Month")
plt.show()
```



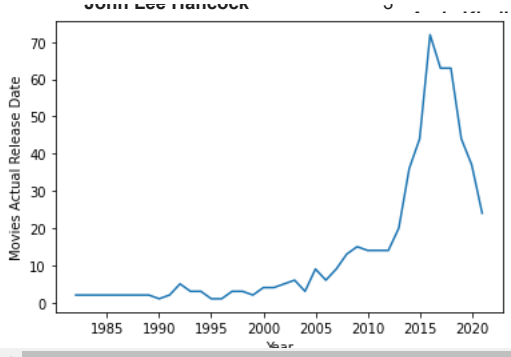
TV Shows are added in Netflix by a tremendous amount in March in UK

Movies are added in Netflix in India by a tremendous amount in first week/last month of current year and first month of next year

```
df_release_year=df_uk_shows[df_uk_shows['release_year']>=1980].groupby(['release_year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_release_year, x='release_year', y='title')
plt.ylabel("Movies Actual Release Date")
plt.xlabel("Year")
plt.show()
```




```
df_release_year=df_uk_movies[df_uk_movies['release_year']>=1980].groupby(['release_year']).agg({"title":"nunique"}).reset_index()
sns.lineplot(data=df_release_year, x='release_year', y='title')
plt.ylabel("Movies Actual Release Date")
plt.xlabel("Year")
plt.show()
```



Same trend of reduction in movies and shows after 2020.

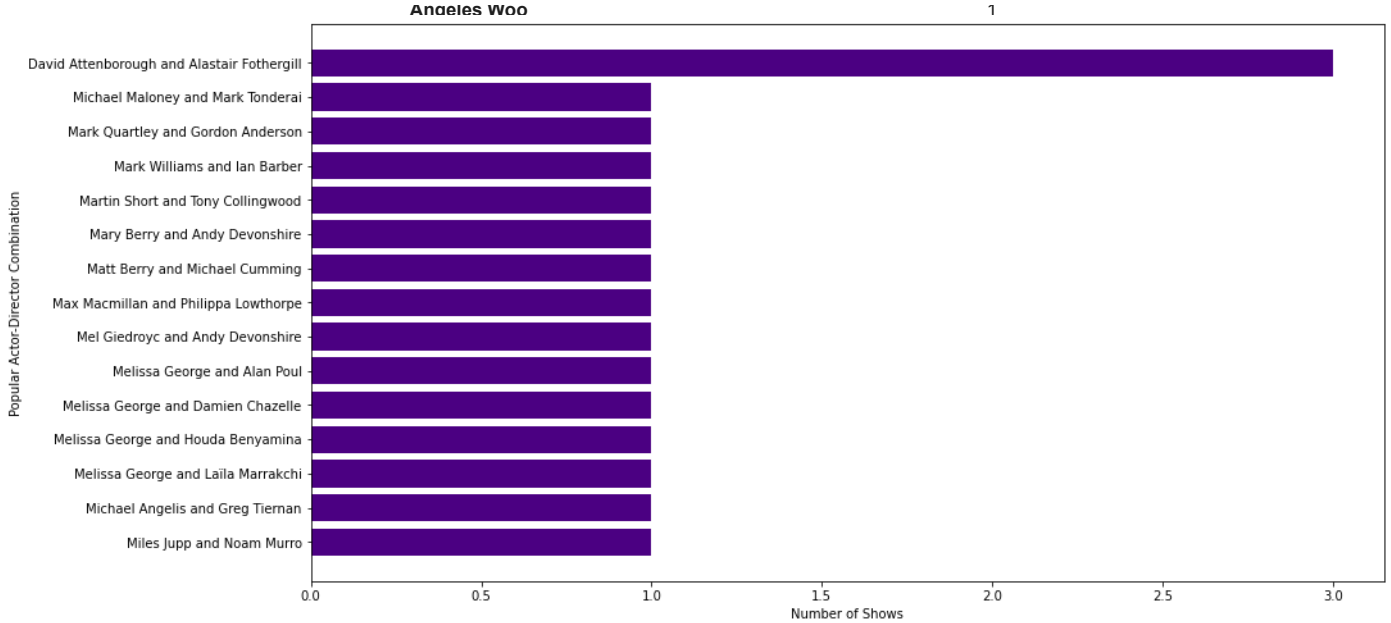
```
#Analysing a combination of actors and directors
df_uk_movies['Actor_Director_Combination'] = df_uk_movies.Actors.str.cat(df_uk_movies.Directors, sep=' and ')
df_uk_movies_subset=df_uk_movies[df_uk_movies['Actors']!='Unknown Actor']
df_uk_movies_subset=df_uk_movies_subset[df_uk_movies_subset['Directors']!='Unknown Director']
df_uk_movies_subset.head()
```

	title	Actors	Directors	genre	country	show_id	type	date_added	release_year	rating	duration	Modified_Added_date
182	Sankofa	Kofi Ghanaba	Haile Gerima	Dramas	United Kingdom	s8	Movie	September 24, 2021	1993	TV-MA	120-150	2021-09-24
188	Sankofa	Kofi Ghanaba	Haile Gerima	Independent Movies	United Kingdom	s8	Movie	September 24, 2021	1993	TV-MA	120-150	2021-09-24
194	Sankofa	Kofi Ghanaba	Haile Gerima	International Movies	United Kingdom	s8	Movie	September 24, 2021	1993	TV-MA	120-150	2021-09-24
200	Sankofa	Oyafunmike Ogunlano	Haile Gerima	Dramas	United Kingdom	s8	Movie	September 24, 2021	1993	TV-MA	120-150	2021-09-24
206	Sankofa	Oyafunmike Ogunlano	Haile Gerima	Independent Movies	United Kingdom	s8	Movie	September 24, 2021	1993	TV-MA	120-150	2021-09-24

```
df_uk_shows['Actor_Director_Combination'] = df_uk_shows.Actors.str.cat(df_uk_shows.Directors, sep=' and ')
df_uk_shows_subset=df_uk_shows[df_uk_shows['Actors']!='Unknown Actor']
df_uk_shows_subset=df_uk_shows_subset[df_uk_shows_subset['Directors']!='Unknown Director']
df_uk_shows_subset.head()
```

	title	Actors	Directors	Genre	country	show_id	type	date_added	release_year	rating	duration	Modified_Added_date	month_a
323	The Great British Baking Show	Mel Giedroyc	Andy Devonshire	British TV Shows	United Kingdom	s9	TV Show	September 24, 2021	2021	TV-14	9 Seasons	2021-09-24	
324	The Great British Baking Show	Mel Giedroyc	Andy Devonshire	Reality TV	United Kingdom	s9	TV Show	September 24, 2021	2021	TV-14	9 Seasons	2021-09-24	
325	The Great British Baking Show	Sue Perkins	Andy Devonshire	British TV Shows	United Kingdom	s9	TV Show	September 24, 2021	2021	TV-14	9 Seasons	2021-09-24	
326	The Great British Baking Show	Sue Perkins	Andy Devonshire	Reality TV	United Kingdom	s9	TV Show	September 24, 2021	2021	TV-14	9 Seasons	2021-09-24	
327	The Great British Baking Show	Mary Berry	Andy Devonshire	British TV Shows	United Kingdom	s9	TV Show	September 24, 2021	2021	TV-14	9 Seasons	2021-09-24	

```
df_actors_directors=df_uk_shows_subset.groupby(['Actor_Director_Combination']).agg({"title":"nunique"}).reset_index().sort_values(by=['title
plt.figure(figsize=(15,8))
plt.barh(df_actors_directors[::-1]['Actor_Director_Combination'], df_actors_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Shows')
plt.ylabel('Popular Actor-Director Combination')
plt.show()
```




```
df_actors_directors=df_uk_movies_subset.groupby(['Actor_Director_Combination']).agg({"title":"nunique"}).reset_index().sort_values(by=['titl
plt.figure(figsize=(15,8))
plt.barh(df_actors_directors[::-1]['Actor_Director_Combination'], df_actors_directors[::-1]['title'],color=['indigo'])
plt.xlabel('Number of Movies')
plt.ylabel('Popular Actor-Director Combination')
plt.show()
```