

NETFLIX DATA ANALYSIS

Importing the required Libraries

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
import seaborn as sns
```

Reading the Dataset

```
In [2]: nf=pd.read_csv('C:/Users/Acer/Desktop/PROJECT HARSH/DATASET/dataset.nf.csv')
```

```
In [3]: nf.head(5)
```

Out[3]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	Int T TV
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	Int
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 min	Int
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	80 min	Inc M
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	

```
In [4]: nf.tail(5)
```

Out[4]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration
7784	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car...	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	TV-MA	99 mir
7785	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	TV-14	111 mir
7786	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	TV-MA	44 mir
7787	s7786	TV Show	Zumbo's Just Desserts	NaN	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	TV-PG	1 Season
7788	s7787	Movie	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	Sam Dunn	NaN	United Kingdom, Canada, United States	March 1, 2020	TV-MA	90 mir

Information about Dataset

In [5]: `nf.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7789 entries, 0 to 7788
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Show_Id         7789 non-null   object
1   Category        7789 non-null   object
2   Title           7789 non-null   object
3   Director        5401 non-null   object
4   Cast            7071 non-null   object
5   Country         7282 non-null   object
6   Release_Date    7779 non-null   object
7   Rating          7782 non-null   object
8   Duration        7789 non-null   object
9   Type            7789 non-null   object
10  Description     7789 non-null   object
dtypes: object(11)
memory usage: 669.5+ KB
```

Describing the Dataset

In [6]: `nf.describe()`

Out[6]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Dur
count	7789	7789	7789	5401	7071	7282	7779	7782	
unique	7787	2	7787	4050	6831	681	1565	14	
top	s6621	Movie	The Lost Okoroshi	Raúl Campos, Jan Suter	David Attenborough	United States	January 1, 2020	TV-MA	1 S
freq	2	5379	2	18	18	2556	118	2865	

Shape of DATASET

In [7]:

nf.shape

Out[7]:

(7789, 11)

Columns of Dataset

In [8]:

nf.columns

Out[8]:

Index(['Show_Id', 'Category', 'Title', 'Director', 'Cast', 'Country', 'Release_Date', 'Rating', 'Duration', 'Type', 'Description'], dtype='object')

Checking Duplicate Value

In [9]:

nf.duplicated().sum()

Out[9]:

2

In [10]:

nf[nf.duplicated()]

Out[10]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration
6300	s684	Movie	Backfire	Dave Patten	Black Deniro, Byron "Squally" Vinson, Dominic ...	United States	April 5, 2019	TV-MA	97 min
6622	s6621	Movie	The Lost Okoroshi	Abba T. Makama	Seun Ajayi, Judith Audu, Tope Tedela, Ifu Enna...	Nigeria	September 4, 2020	TV-MA	94 min

In [11]:

nf.drop_duplicates(inplace=True)

```
In [12]: nf[nf.duplicated()]
```

```
Out[12]:
```

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Descr
---------	----------	-------	----------	------	---------	--------------	--------	----------	------	-------

DROPPED all Duplicated values

Checking unique values

```
In [13]: nf.nunique()
```

```
Out[13]:
```

Show_Id	7787
Category	2
Title	7787
Director	4050
Cast	6831
Country	681
Release_Date	1565
Rating	14
Duration	216
Type	492
Description	7769

dtype: int64

Checking Null-Values

```
In [14]: nf.isnull().sum()
```

```
Out[14]:
```

Show_Id	0
Category	0
Title	0
Director	2388
Cast	718
Country	507
Release_Date	10
Rating	7
Duration	0
Type	0
Description	0

dtype: int64

DROP NULL VALUES

```
In [15]: nf.dropna()
```

Out[15]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Dur
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	9
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	7
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	8
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	12
5	s6	TV Show	46	Serdar Akar	Erdal Beşikçioğlu, Yasemin Allen, Melis Birkan...	Turkey	July 1, 2017	TV-MA	1 S
...
7780	s7779	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	R	8
7782	s7781	Movie	Zoo	Shlok Sharma	Shashank Arora, Shweta Tripathi, Rahul Kumar, ...	India	July 1, 2018	TV-MA	9
7783	s7782	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	PG	8
7784	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car...	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	TV-MA	9
7785	s7784	Movie	Zubaan	Mozes Singh	Vicky Kaushal,	India	March 2, 2019	TV-14	11

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Dur
				Sarah-Jane Dias, Raaghav Chanan...				
1000	11							

```
In [16]: nf.shape
```

Out[16]: (7787, 11)

Datatypes

```
In [17]: nf.dtypes
```

Out[17]: Show_Id object
Category object
Title object
Director object
Cast object
Country object
Release_Date object
Rating object
Duration object
Type object
Description object
dtype: object

CONVERTING DATE AND TIME FORMAT

```
In [18]: #adding data columns  
nf['Date']=pd.to_datetime(nf['Release_Date'])  
#adding Year columns  
nf['Year']=nf['Date'].dt.year
```

```
In [19]: nf.head()
```

Out[19]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	Int T TV
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	Int
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 min	Int
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	80 min	Inc M
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	

FOR FINDING ANY PARTICULAR MOVIE AND ALL RECORDS ABOUT THAT.

In [20]:

```
nf[nf['Title'].isin(['Billu'])]
```

Out[20]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	
926	s927	Movie	Billu	Priyadarshan	Irrfan Khan, Shah Rukh Khan, Lara Dutta, Rajpa...	India	April 1, 2018	TV-PG	132 min	Int

TOP 10 OLDEST TV SHOWS & MOVIES IN NETFLIX DATASET

In [21]:

```
#oldest movies
Om=nf[nf['Category'] == 'Movie'].sort_values('Year').head(20)
Om[['Title','Release_Date','Rating','Cast']]
```

Out[21]:

	Title	Release_Date	Rating	Cast
7115	To and From New York	January 1, 2008	TV-MA	Barbara King, Shaana Diya, John Krikukenas, Y...
3248	Just Another Love Story	May 5, 2009	TV-MA	Anders W. Berthelsen, Rebecka Hemse, Nikolaj L...
5766	Splatter	November 18, 2009	TV-MA	Corey Feldman, Tony Todd, Tara Leigh, Erin Way...
3840	Mad Ron's Prevues from Hell	November 1, 2010	NR	Nick Pawlow, Jordy Schell, Jay Kushwara, Micha...
2041	Even the Rain	May 17, 2011	TV-MA	Luis Tosar, Gael García Bernal, Juan Carlos Ad...
7151	Too Young the Hero	October 1, 2011	TV-MA	Ricky Schroder, Jon DeVries, Debra Mooney, Mar...
2962	In Defense of a Married Man	October 1, 2011	TV-14	Judith Light, Michael Ontkian, Jerry Orbach, P...
7232	Triumph of the Heart	October 1, 2011	TV-PG	Mario Van Peebles, Susan Ruttan, Lane R. Davis...
7364	Unspeakable Acts	October 1, 2011	TV-14	Jill Clayburgh, Brad Davis, Sam Behrens
7416	Victim of Beauty	October 1, 2011	TV-14	William Devane, Jeri Ryan, Michele Abrams, Nic...
6847	The Ryan White Story	October 1, 2011	TV-PG	Judith Light, Lukas Haas, Michael Bowen, Nikki...
232	A Stoning in Fulham County	October 1, 2011	TV-14	Ken Olin, Jill Eikenberry, Maureen Mueller, Gr...
5833	Strange Voices	October 1, 2011	TV-PG	Nancy McKeon, Valerie Harper, Stephen Macht, T...
308	Adam: His Song Continues	October 1, 2011	TV-MA	Daniel J. Travanti, JoBeth Williams, Richard M...
2622	Hard Lessons	October 1, 2011	TV-14	Denzel Washington, Lynn Whitfield, Richard Mas...
3221	Joseph: King of Dreams	September 27, 2011	TV-PG	Ben Affleck, Mark Hamill, Richard Herd, Mauree...
5042	Quiet Victory: The Charlie Wedemeyer Story	October 1, 2011	TV-PG	Pam Dawber, Michael Nouri, Bess Meyer, Peter B...
1229	Casa de mi Padre	November 14, 2012	R	Will Ferrell, Gael García Bernal, Diego Luna, ...
3466	Kung Fu Panda: Holiday	December 1, 2012	TV-PG	Jack Black, Angelina Jolie, Dustin Hoffman, Ja...
818	Being Elmo: A Puppeteer's Journey	February 21, 2012	PG	Kevin Clash, Whoopi Goldberg

In [22]:

```
#oldest TV Shows
Os=nf[nf['Category'] == 'TV Show'].sort_values('Year').head(20)
Os[['Title', 'Release_Date', 'Rating', 'Cast']]
```


Out[22]:

	Title	Release_Date	Rating	Cast
1765	Dinner for Five	February 4, 2008	TV-MA	NaN
2503	Gossip Girl	October 8, 2013	TV-14	Blake Lively, Leighton Meester, Penn Badgley, ...
6057	The 4400	September 1, 2013	TV-14	Joel Gretsch, Jacqueline McKenzie, Patrick Joh...
1089	Breaking Bad	August 2, 2013	TV-MA	Bryan Cranston, Aaron Paul, Anna Gunn, Dean No...
3098	Jack Taylor	March 31, 2013	TV-MA	Iain Glen, Killian Scott, Nora-Jane Noone, Tar...
5313	Russell Peters vs. the World	October 14, 2013	NR	Russell Peters
6631	The Magic School Bus	November 19, 2014	TV-Y	Lily Tomlin, Daniel DeSanto, Erica Luttrell, M...
2498	Goosebumps	November 19, 2014	TV-G	R.L. Stine
4811	Pee-wee's Playhouse	December 18, 2014	TV-PG	Paul Reubens
6158	The Borgias	February 1, 2014	TV-MA	Jeremy Irons, François Arnaud, Holliday Graing...
3643	Lilyhammer	November 21, 2014	TV-MA	Steven Van Zandt, Trond Fausa, Steinar Sagen, ...
7522	Weeds	April 1, 2014	TV-MA	Mary-Louise Parker, Hunter Parrish, Alexander ...
4685	Oscar's Oasis	March 15, 2015	TV-Y	NaN
5742	Spartacus	February 1, 2015	TV-MA	Andy Whitfield, Liam McIntyre, Dustin Clare, J...
5202	Richie Rich	May 22, 2015	TV-G	Jake Brennan, Joshua Carlon, Lauren Taylor, Je...
2658	Haven	September 8, 2015	TV-14	Emily Rose, Lucas Bryant, Nicholas Campbell, E...
5792	Star-Crossed	April 16, 2015	TV-14	Aimee Teegarden, Matt Lanter, Grey Damon, Male...
6275	The Day My Butt Went Psycho!	June 30, 2015	TV-Y7	Mac Heywood, Robert Tinkler, Bryn McAuley, Ked...
766	Battle Creek	December 31, 2015	TV-14	Josh Duhamel, Dean Winters, Aubrey Dollar, Edw...
7544	Wet Hot American Summer	July 31, 2015	TV-MA	Elizabeth Banks, Lake Bell, H. Jon Benjamin, M...

In [23]: `nf[(nf['Category']=='Movie')&(nf['Country']=='India')]['Title'].tail(10)`

```
Out[23]: 7704          Yeh Hai Bakrapur
7746      Yours Sincerely, Kanan Gill
7756          Yuva
7764          Zed Plus
7765          Zenda
7766          Zero
7771      Zindagi Na Milegi Dobara
7778          Zokkomon
7782          Zoo
7785          Zubaan
Name: Title, dtype: object
```

Indian MOVIES Released on NETFLIX in 2016

```
In [24]: nf[(nf['Category']=='Movie')&(nf['Country']=='India')&(nf['Year']==2016)][['Title']]

Out[24]: 59          1000 Rupee Note
1734          Dhanak
2107      Famous in Ahmedabad
2639          Harud
3027      Interrogation
3200          John & Jane
4276      Mumbai Cha Raja
6547      The Land of Cards
6983      The Violin Player
7312          Umrika
7686      X: Past Is Present
Name: Title, dtype: object
```

TOP 10 CAST OF MOVIES IN NETFLIX BASED ON NUMBER OF MOVIES

```
In [25]: k=nf[(nf['Category']=='Movie')][['Cast']]

In [26]: k.value_counts().head(10)

Out[26]: Samuel West
10
Jeff Dunham
7
Kevin Hart
6
Craig Sechler
6
Michela Luci, Jamie Watson, Eric Peterson, Anna Claire Bartlam, Nicolas Aquí, Cory
Doran, Julie Lemieux, Derek McGrath      5
Jim Gaffigan
5
Bill Burr
5
Iliza Shlesinger
5
David Attenborough
5
Mike Birbiglia
4
Name: Cast, dtype: int64
```

TOP 10 CAST OF TV SHOWS IN NETFLIX BASED ON NUMBER OF TV SHOWS

```
In [27]: s=nf[(nf['Category']=='TV Show')][['Cast']]
```

```
In [28]: s.value_counts().head(10)
```

```
Out[28]: David Attenborough
13
Michela Luci, Jamie Watson, Anna Claire Bartlam, Dante Zee, Eric Peterson
4
Bettany Hughes
2
Dave Chappelle
2
Bob Ross
2
Phil Vischer, Mike Nawrocki, Rob Paulsen, Tress MacNeille
2
Brennan Mejia, Camille Hyde, Yoshi Sudarso, Michael Taber, James Davies, Claire Blackwelder
2
Chris Packham
2
You, Reina Triendl, Yoshimi Tokui, Azusa Babazono, Ryota Yamasato, Shono Hayama
2
Anitta
2
Name: Cast, dtype: int64
```

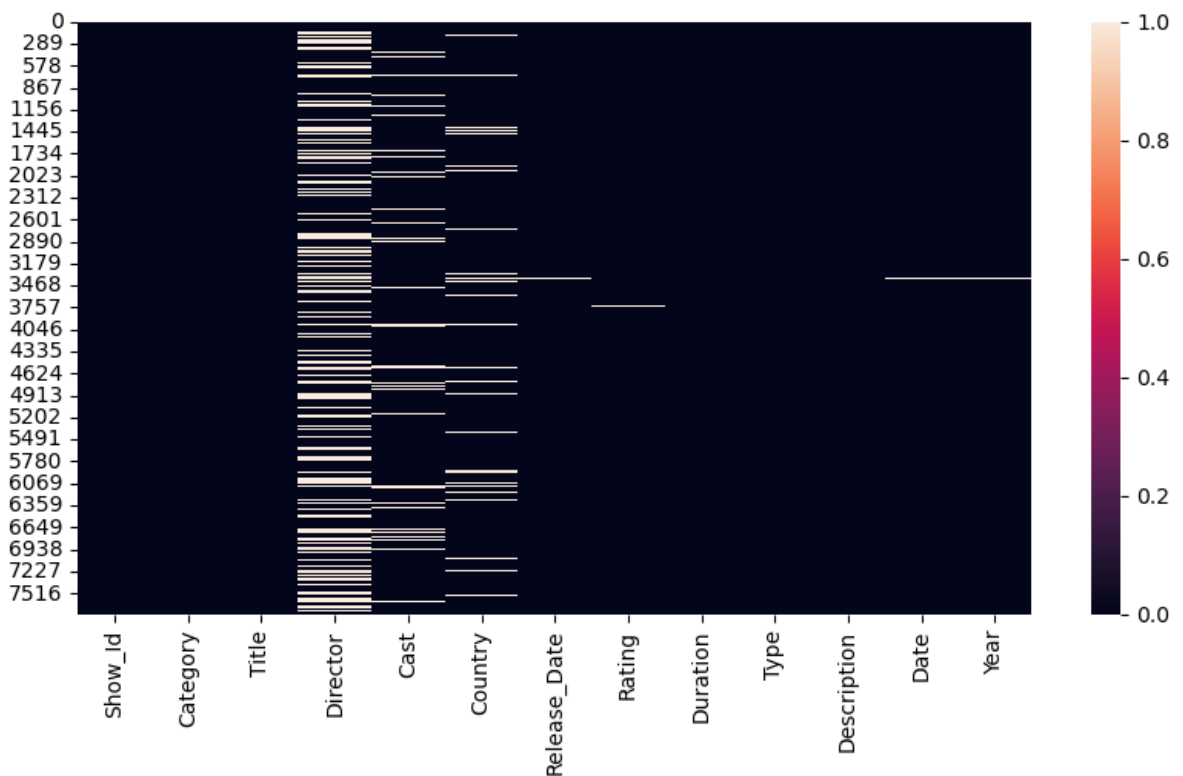
```
In [ ]:
```

DATA VISUALIZATION

Using Heatmap to showing Null-Values

```
In [29]: plt.figure(figsize=[10,5])
sns.heatmap(nf.isnull())
```

```
Out[29]: <AxesSubplot:>
```

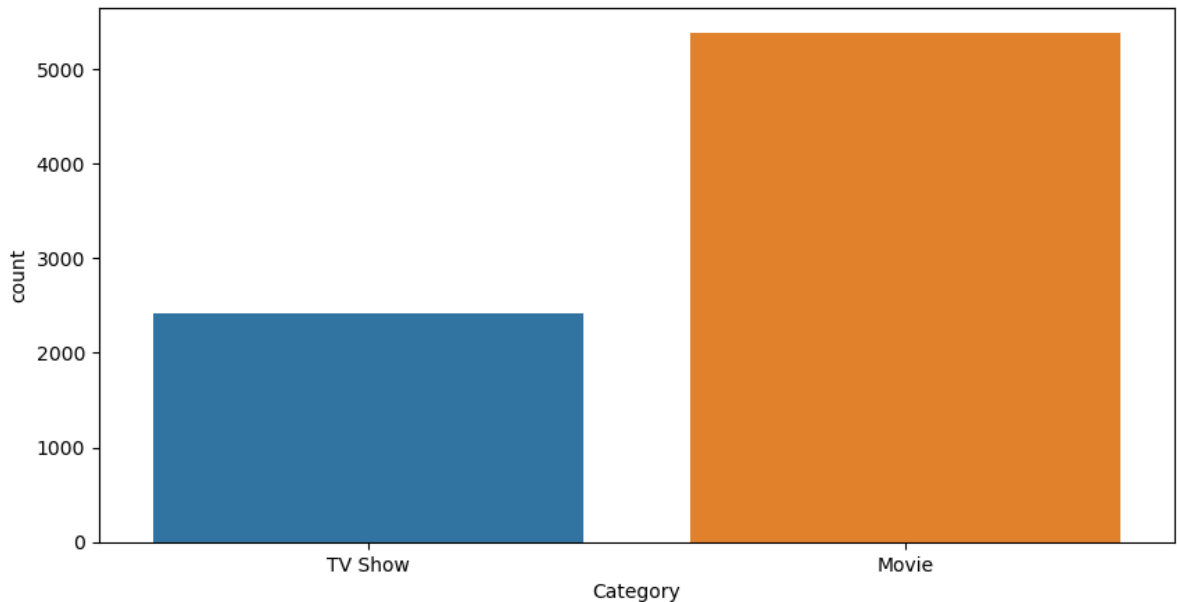


```
In [30]: #MOVIES AND TV-SHOWS IN A DATASET  
nf['Category'].value_counts()
```

```
Out[30]: Movie      5377  
TV Show    2410  
Name: Category, dtype: int64
```

```
In [31]: plt.figure(figsize=[10,5])  
sns.countplot(x='Category', data=nf)
```

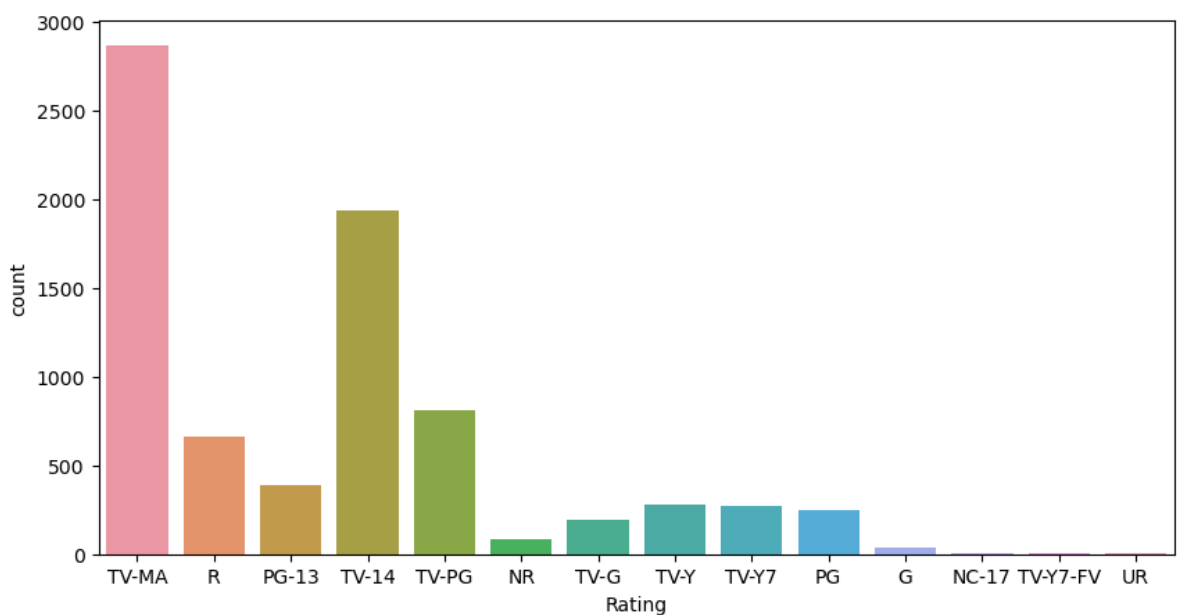
```
Out[31]: <AxesSubplot:xlabel='Category', ylabel='count'>
```



RATINGS(MOVIES & TV SHOWS)

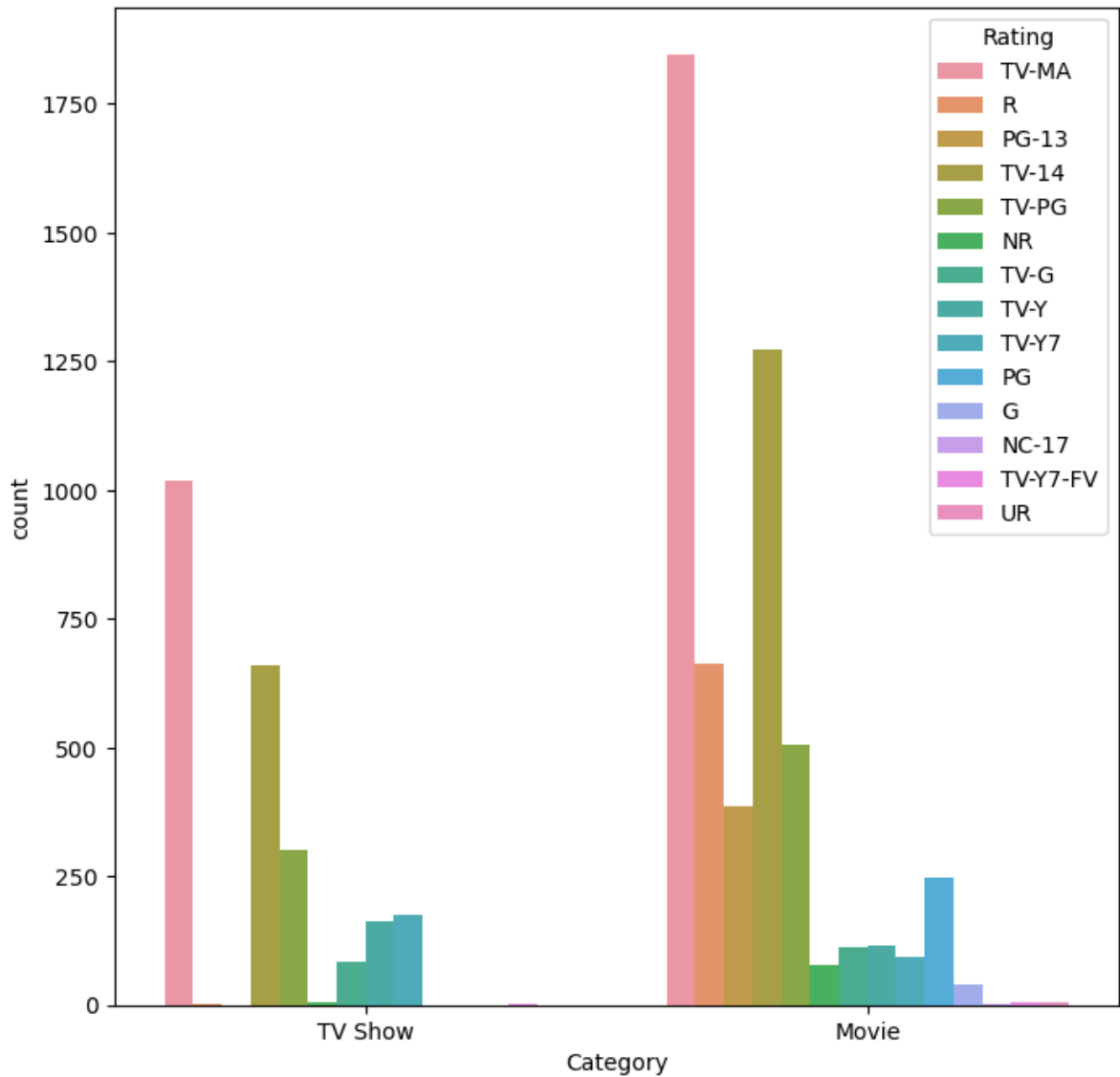
```
In [32]: plt.figure(figsize=[10,5])  
sns.countplot(x='Rating', data=nf)
```

```
Out[32]: <AxesSubplot:xlabel='Rating', ylabel='count'>
```



TV-SHOWS BY RATINGS & MOVIES BY RATINGS

```
In [33]: plt.figure(figsize=(8,8))
sns.countplot(x='Category',hue='Rating',data=nf)
plt.show()
```

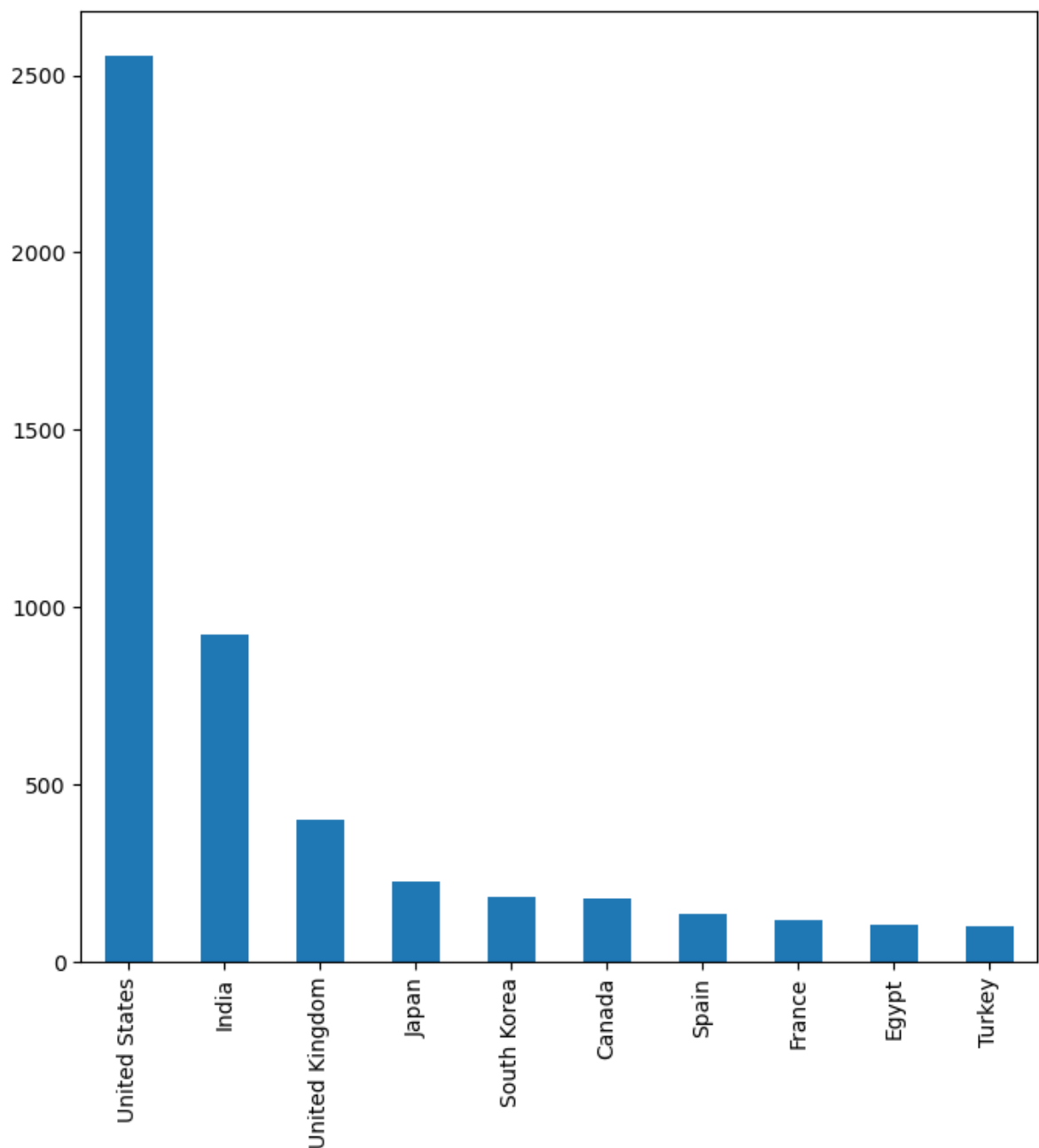


TOP 10 COUNTRIES ON NETFLIX

```
In [34]: C=nf.Country.value_counts().head(10)
```

```
In [35]: plt.figure(figsize=[8,8])
C.plot(kind='bar')
```

```
Out[35]: <AxesSubplot:>
```

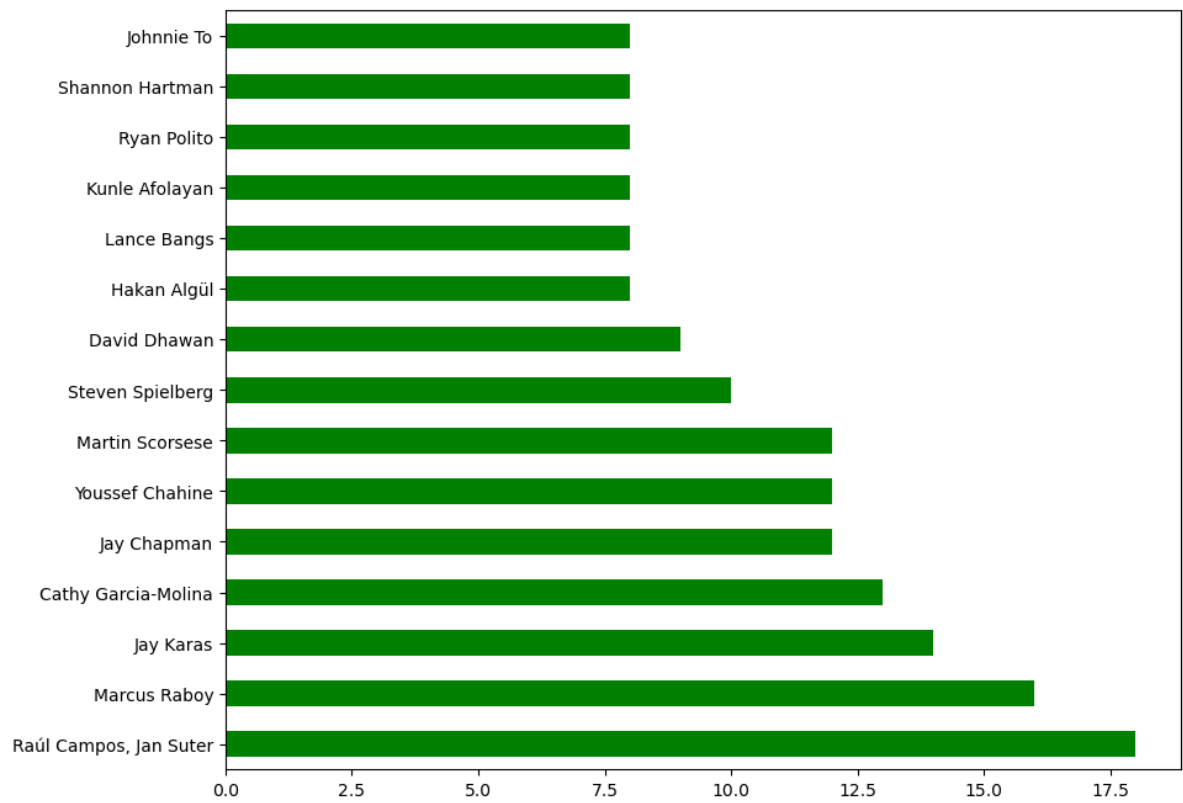


TOP 15 DIRECTORS ON NETFLIX

```
In [36]: T15D=nf.Director.value_counts().head(15)
```

```
In [37]: plt.figure(figsize=[10,8])  
T15D.plot(kind='barh',color='Green')
```

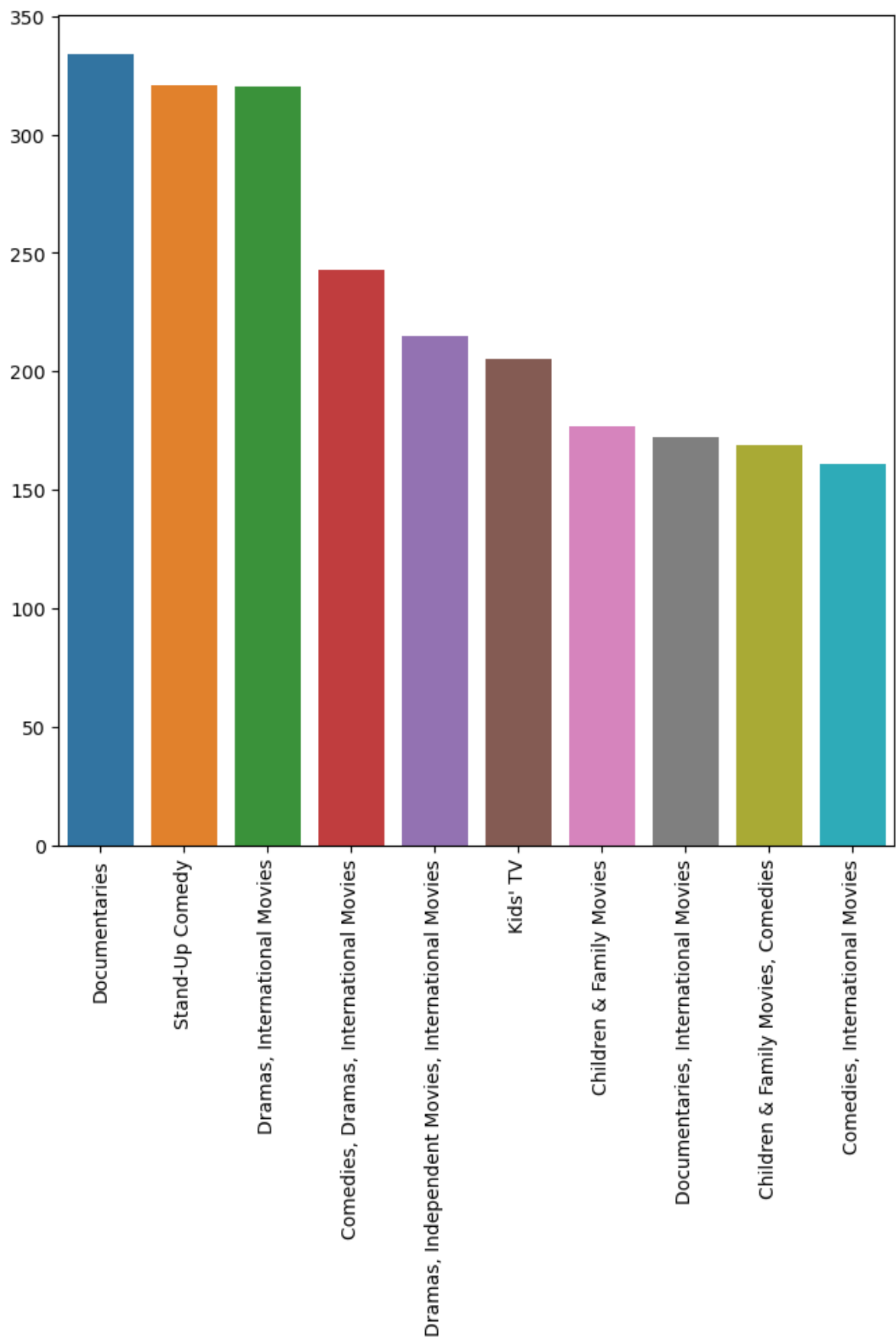
```
Out[37]: <AxesSubplot:>
```



Top 10 Generes in Netflix

```
In [38]: Types=nf['Type'].value_counts().head(10)
```

```
In [39]: plt.figure(figsize=(8,8))
sns.barplot(y=Types.values, x=Types.index)
plt.xticks(rotation = 90)
plt.show()
```

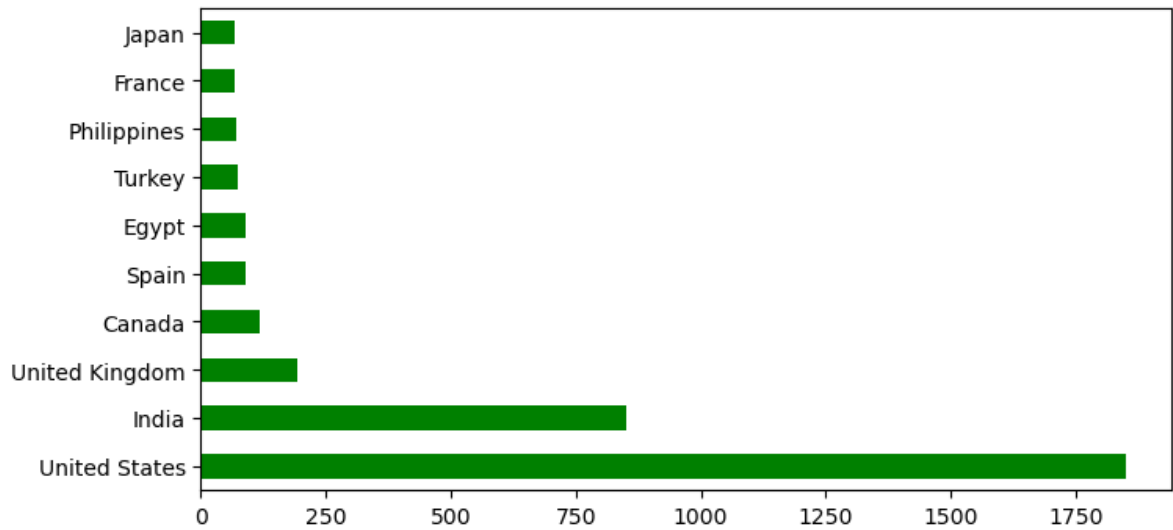


TOP 10 COUNTRIES MOST MOVIES RELEASED BY COUNTRY

```
In [40]: MOSTMOVIES=nf[nf['Category']=='Movie']  
mm=MOSTMOVIES.Country.value_counts().head(10)
```

```
In [41]: plt.figure(figsize=[8,4])  
mm.plot(kind="barh",color='green')
```


Out[41]: <AxesSubplot:>

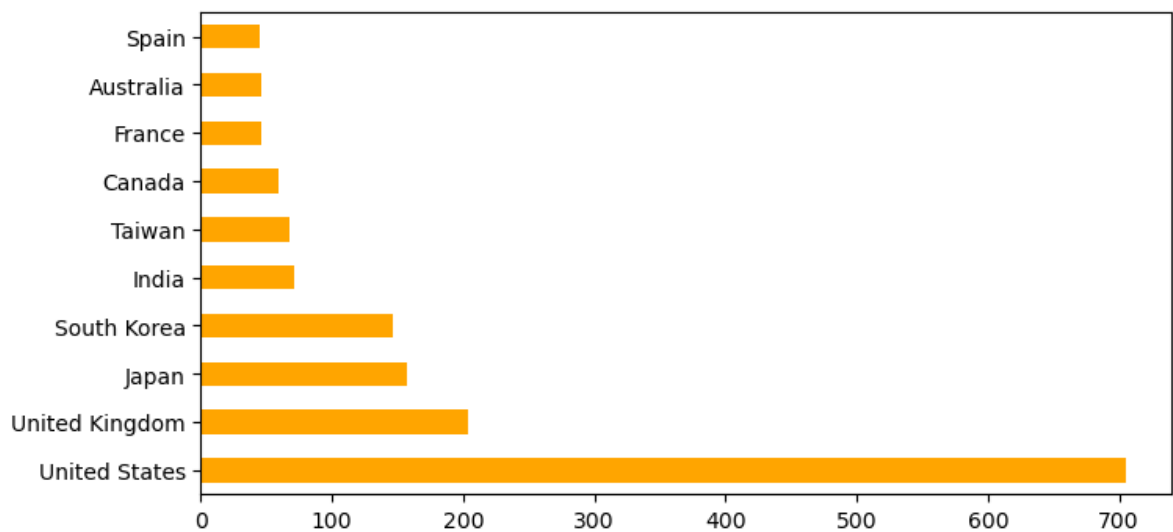


TOP 10 COUNTRIES MOST TV SHOWS RELEASED BY COUNTRY

```
In [42]: MOSTTV=nf[nf['Category']=='TV Show']
TV1=MOSTTV.Country.value_counts().head(10)
```

```
In [43]: plt.figure(figsize=[8,4])
TV1.plot(kind="barh",color='Orange')
```

Out[43]: <AxesSubplot:>

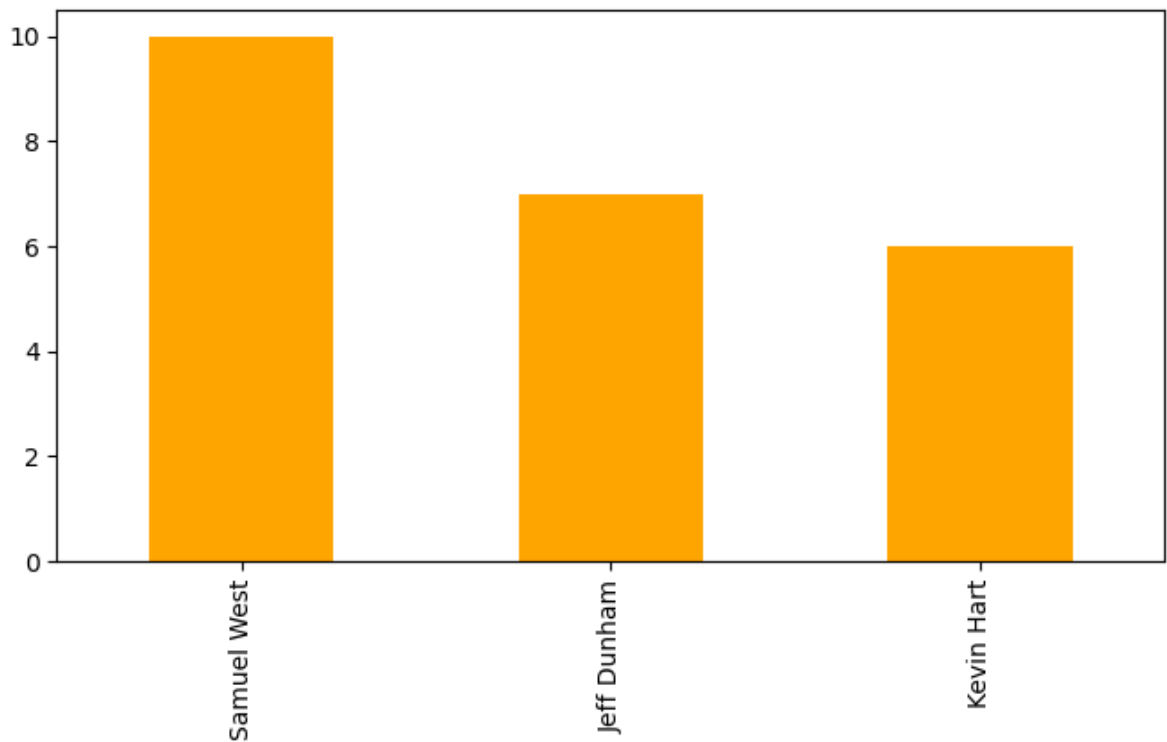


TOP 3 ACTORS OF Movies

```
In [44]: k=nf[(nf['Category']=='Movie')]['Cast']
kk=k.value_counts().head(3)
```

```
In [45]: plt.figure(figsize=[8,4])
kk.plot(kind="bar",color='Orange')
```

Out[45]: <AxesSubplot:>

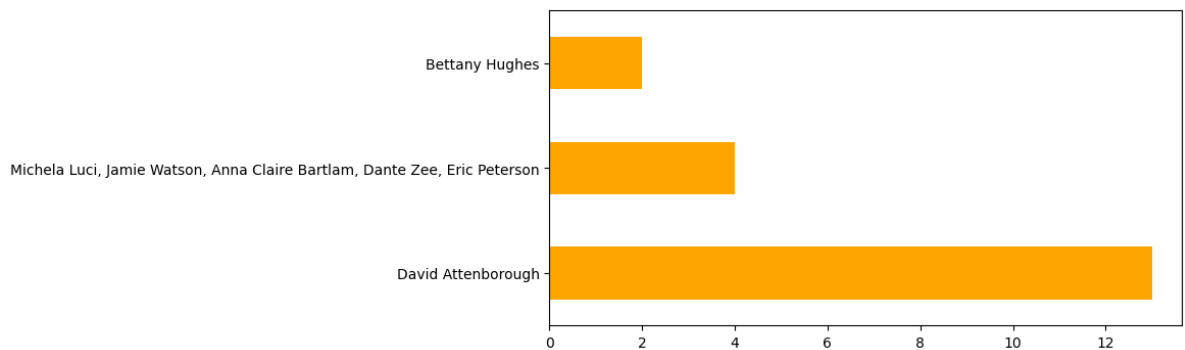


TOP 3 ACTORS OF TV SHOWS

```
In [46]: s=nf[(nf['Category']=='TV Show')]['Cast']
ss=s.value_counts().head(3)
```

```
In [47]: plt.figure(figsize=[8,4])
ss.plot(kind="barh",color='Orange')
```

Out[47]: <AxesSubplot:>

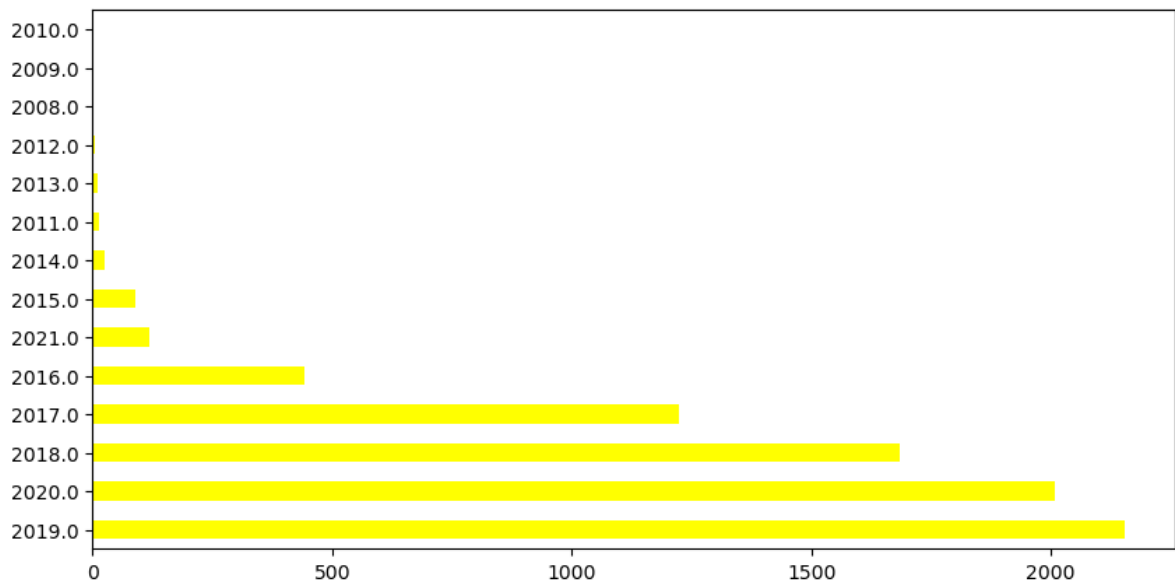


Number of shows released per year

```
In [48]: YS=nf['Year'].value_counts()
```

```
In [49]: plt.figure(figsize=[10,5])
YS.plot(kind='barh',color='yellow')
```

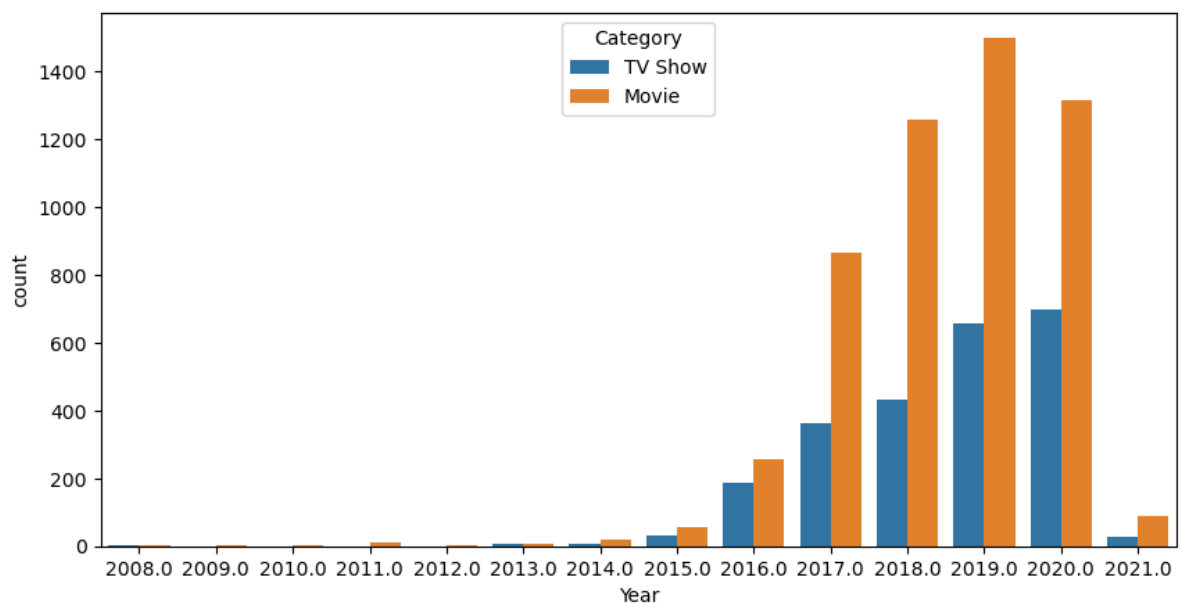
Out[49]: <AxesSubplot:>



Comparison Between TV SHOWS & MOVIES respect to Year

```
In [50]: plt.figure(figsize=[10,5])
sns.countplot(data=nf,x='Year',hue='Category')
```

```
Out[50]: <AxesSubplot:xlabel='Year', ylabel='count'>
```



Duration of TV-Shows & Movies

```
In [51]: plt.rcParams['figure.figsize'] = (16,8)
sns.countplot(y = 'Duration', data = nf, order = nf['Duration'].value_counts()[:10])
plt.title('DURATION ', fontsize = 20)
```

```
Out[51]: Text(0.5, 1.0, 'DURATION ')
```

