

This Netflix Dataset has information about the TV Shows and Movies available on Netflix till 2021. This dataset is collected from Flixable which is a Third-Party Netflix search engine, and available on Kaggle website for free.

```
import pandas as pd

df=pd.read_csv(r'/content/8. Netflix Dataset.csv')

df
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	De
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	wh
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	Dramas, International Movies	earl
			23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 min	Horror Movies, International Movies	Wh rec
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	80 min	Action & Adventure, Independent Movies, Sci-Fi...	posi wc
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	Dramas	A br be
...	...	...	...	...	...	...	...	...	...	...	...
7784	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car... Vickv	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	TV-MA	99 min	Dramas, International Movies	Leb V Zc

```
df.head()
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Descriptio
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a futu where the eli inhabit island
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	Dramas, International Movies	After devastatir earthquake hi Mexico Cit
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, ...	Singapore	December 20, 2018	R	78 min	Horror Movies, International	When an am recruit is four dead, h

```
df.tail()
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Descr
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 min	Dramas, International Movies	Le C c Zozo i
7785	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	TV-14	111 min	Dramas, International Movies, Music & Musicals	A: but p wc we
7786	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	TV-MA	44 min	Documentaries, International Movies, Music & M...	docun South

```
df.shape
```

(7789, 11)

```
df.describe()
```

Saved successfully!

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	De
count	7789	7789	7789	5401	7071	7282	7779	7782	7789	7789	
unique	7787	2	7787	4050	6831	681	1565	14	216	492	
top	s6621	Movie	The Lost Okoroshi	Raúl Campos, Jan Suter	David Attenborough	United States	January 1, 2020	TV-MA	1 Season	Documentaries	h
freq	2	5379	2	18	18	2556	118	2865	1608	334	

```
df.size
```

85679

```
df.columns
```

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

```
df.dtypes
```

```
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
```

```
df.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
```

1.Is there any Duplicate records in this dataset? If yes, then remove the duplicate records.

```
df.shape # we have Rows= 7789 and Columns=11

(7789, 11)
```

```
df[df.duplicated()] #lets see the duplicated
```

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	
6300	s684	Movie	Backfire	Dave Patten	Black Deniro, Byron "Squally" Vinson, Dominic ... Seun Aisavi	United States	April 5, 2019	TV-MA	97 min	Dramas, Independent Movies, Thrillers	When tv would-I robbe accidenta kill a

Saved successfully!

✕ Making permanent removal of duplicates

```
df[df.duplicated()] #so all duplicates removed
```

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



2.Is there any Null Values Present in any Columns? Show with Heat -map

```
df.isnull() #check for Null values
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description
0	False	False	False	True	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...
7784	False	False	False	False	False	False	False	False	False	False	False
7785	False	False	False	False	False	False	False	False	False	False	False
7786	False	False	False	True	False	True	False	False	False	False	False
7787	False	False	False	True	False	False	False	False	False	False	False
7788	False	False	False	False	True	False	False	False	False	False	False

7787 rows × 11 columns

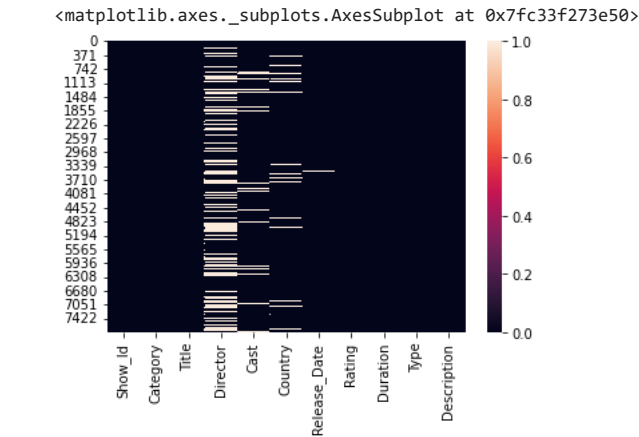
```
df.isnull().sum() #Get a count of total Null values in each Columns
```

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
```

```
#importing Seaborn library for Heat-Map
import seaborn as sns
```

```
sns.heatmap(df.isnull())
```



3.For 'House of cards', what is the Show ID and Who is the Director od this Show?

Saved successfully! ✕

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit a island
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	Dramas, International Movies	After devastating earthquake hits Mexico City
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii,	Singapore	December 20, 2018	R	78 min	Horror Movies, International	When an army recruit is found dead, his

```
#to show all records of a particular item in any column
df[df['Title'].isin(['House of Cards'])]
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	
	2832	s2833	TV Show	House of Cards	Robin Wright, David Fincher, Gerald McRaney,	Kevin Spacey, Robin Wright, Kate Mara, Corey Stoll	United States	November 2, 2018	TV-MA	6 Seasons	TV Dramas, TV Thrillers	A ruthless politician will stop at nothing to ...

```
df[df['Title'].str.contains('House of Cards')]#TO SHOW ALL DETAILS OF A PARTICULAR STRING IN ANY COLUMN
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	
	2832	s2833	TV Show	House of Cards	Robin Wright, David Fincher, Gerald McRaney,	Kevin Spacey, Robin Wright, Kate Mara, Corey Stoll	United States	November 2, 2018	TV-MA	6 Seasons	TV Dramas, TV Thrillers	A ruthless politician will stop at nothing to ...

4. IN WHICH YEAR HIGHEST NUMBERS OF THE TV SHOWS & MOVIES WERE RELEASED ? SHOW WITH BAR GRAPH

df.dtypes

```
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
```

```
df['Date_N']=pd.to_datetime(df['Release_Date'])
```

df.head()

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a futur where the eli inhabit ε island
				Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	Dramas, International Movies	After devastatir earthquake hi Mexico Cit
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 min	Horror Movies, International Movies	When an arm recruit is four dead, h fellow

df.dtypes

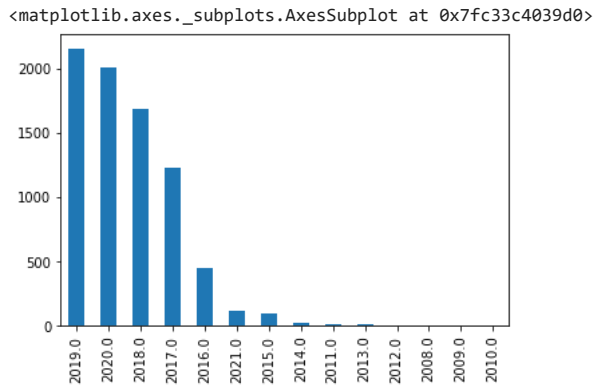
```
Show_Id      object
Category     object
Title        object
Director     object
Cast         object
Country      object
Release_Date  object
Rating       object
Duration     object
Type         object
Description  object
Date_N       datetime64[ns]
dtype: object
```

```
df['Date_N'].dt.year.value_counts()
#it counts the occurance of all individual years in date column
# 2019 has highest number of movies and Tv shows release i.e= 2153
```

```
2019.0    2153
2020.0    2009
2018.0    1685
2017.0    1225
2016.0     443
2021.0     117
2015.0      88
2014.0      25
2011.0      13
2013.0      11
2012.0       3
2008.0       2
2009.0       2
2010.0       1
Name: Date_N, dtype: int64
```

BAR GRAPH

```
df['Date_N'].dt.year.value_counts().plot(kind='bar')
```



5. How many Movies & TV Shows are in the dataset? Show with Bar Graph

```
#GROUP BY: all unique items TV shows and Movies and get a count of all movies and Tv shows
df.groupby('Category').Category.count()
```

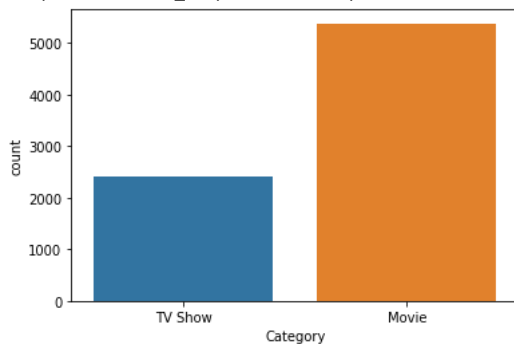
```
Category
Movie      5377
TV Show    2410
Name: Category, dtype: int64
```

Saved successfully!

unique values of any column in the form of a bar graph.

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword argument to .warn():
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fc33c420a90>
```



## ▼ Show all the Movies that are released in year 2020

```
#creating new column only on 'Year'
df['Year']=df['Date_N'].dt.year
df.head(2)
```

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...
				Jorge	Demián Bichir, Héctor	December 22				Dramas,	After a devastating

```
#Filtering the Year 2020 and movies that released in that particular year
df[(df['Category']=='Movie') & (df['Year']==2020)]
```

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	De
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	Dramas
6	s7	Movie	122	Yasir Al Yasiri	Amina Khalil, Ahmed Dawood, Tarek Lotfy, Ahmed...	Egypt	June 1, 2020	TV-MA	95 min	Horror Movies, International Movies
14	s15	Movie	3022	John Suits	Omar Epps, Kate Walsh, Miranda Cosgrove, Angus...	United States	March 19, 2020	R	91 min	Independent Movies, Sci-Fi & Fantasy, Thrillers
27	s28	Movie	#Alive	Cho Il	Yoo Ah-in, Park Shin-hye	South Korea	September 8, 2020	TV-MA	99 min	Horror Movies, International Movies, Thrillers
28	s29	Movie	#AnneFrank - Parallel stories	Sabina Fedeli, Anna Migotto	Helen Mirren, Gengher Gatti	Italy	July 1, 2020	TV-14	95 min	Documentaries, International Movies
...	...	...	...	...	...	...	...	...	...	...
7762	s7761	Movie	Zaki Chan	Wael Ihsan	Ahmed Helmy, Yasmin Abdulaziz, Hassan Hosny, H...	Egypt	May 19, 2020	TV-PG	109 min	Comedies, International Movies, Romantic Movies
7783	s7782	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	PG	88 min	Children & Family Movies, Comedies
				Imad Creidi, Antoinette	Sweden, Czech					

Show Only the titles of all Tv Shows that were released in India Only

Gerni ninguom, 7c

```
#Filtering Category as TVshows and Country as India
df[(df['Category']=='TV Show') & (df['Country']=='India')]
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Desc
86	s87	TV Show	21 Sarfarosh: Saragarhi 1897	NaN	Luke Kenny, Mohit Raina, Mukul Dev	India	December 1, 2018	TV-14	1 Season	International TV Shows, TV Dramas	gre: ;
132	s133	TV Show	7 (Seven)	Nizar Shafi	Rahman, Havish, Regina Cassandra, Nandita Swet...	India	July 30, 2019	TV-14	1 Season	TV Shows	re husl
340	s341	TV Show	Agent Raghav	NaN	Sharad Kelkar, Aahana Kumra, Mahesh Manjrekar,...	India	April 15, 2017	TV-14	1 Season	Crime TV Shows, International TV Shows, TV Act...	t erud ski

Show Top 10 Directors, who gave the Highest Number of tv Shows & Movies to Netflix?

VishalMarch 31KIDS TV, TV

#Value\_counts(): it gives an index with DESCENDING order.. High ---> Low

df['Director'].value\_counts().head(10)

Raúl Campos, Jan Suter18

Marcus Raboy16

Jay Karas14

Cathy Garcia-Molina13

Jay Chapman12

Youssef Chahine12

Saved successfully!

X

Hakan Algül8

Name: Director, dtype: int64

Show all the Records where 'Category is Movie and Type is Comedies' or 'Country is united Kingdom'?

```
df[(df['Category']=='Movie') & (df['Type']=='Comedies') | (df['Country']=='United Kingdom')]
```



Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description
19	s20	Movie	'89	NaN	Lee Dixon, Ian Wright, Paul Merson	United Kingdom	May 16, 2018	TV-PG	87 min	Sports Movies
				Nesta						

In how many Movies/TV Shows Tom Cruise was cast

0	s1	TV Show	3%	NaN	Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...
---	----	---------	----	-----	-------------------------------	--------	-----------------	-------	-----------	---------------------------------------------------	---------------------------------------------------

df.head(2)

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...
					Jorge	Demián Bichir, Hótar Thomas		December 23		Dramas,	After a devastating

# Filtering  
df[df['Cast']=='Tom cruise'] #works only when Cast is Only 'Tom Cruise' alone

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
---------	----------	-------	----------	------	---------	--------------	--------	----------	------	-------------	--------	------

Saved successfully!

df[df['Cast'].str.contains('Tom Cruise')]

# Exception: will throw ERROR ...> Cannot mask with non-boolean array containing NA / NaN values  
# this function cannot work if the Column or Row contains NA/NaN values

ValueError

Traceback (most recent call last)

<ipython-input-91-5af05d9847e7> in <module>

----> 1 df[df['Cast'].str.contains('Tom Cruise')]

1 frames

/usr/local/lib/python3.8/dist-packages/pandas/core/common.py

in is\_bool\_indexer(key)

137 # Don't raise on e.g. ["A", "B", np.nan], see

138 # test\_loc\_getitem\_list\_of\_labels\_categoricalindex\_with\_na

--> 139 raise ValueError(na\_msg)

140 return False

141 return True

ValueError: Cannot mask with non-boolean array containing NA / NaN values

SEARCH STACK OVERFLOW

# NEED TO CREATE A NEW DATAFRAME.  
df\_new=df.dropna() # dropping nulls

df\_new.isnull().sum()

Show_Id	0
Category	0
Title	0
Director	0
Cast	0
Country	0
Release_Date	0
Rating	0
Duration	0
Type	0
Description	0
Date_N	0
Year	0
dtype:	int64

df\_new[df\_new['Cast'].str.contains('Tom Cruise')]

https://colab.research.google.com/drive/1DyWIOx\_3ZAPT16993tjNcPQuwUhWW5nb#scrollTo=qifBT3Lf5TVv&printMode=true

9/12

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Descript
3860	s3861	Movie	Magnolia	Paul Thomas Anderson	Jeremy Blackman, Tom Cruise, Melinda Dillon, A...	United States	January 1, 2020	R	189 min	Dramas, Independent Movies	Thro char hur action, I history ar

Q.9.What are the Different ratings Defined by Netflix?

```
df['Rating'].nunique()

14

df['Rating'].unique()

[array(['TV-MA', 'R', 'PG-13', 'TV-14', 'TV-PG', 'NR', 'TV-G', 'TV-Y', nan,
       'TV-Y7', 'PG', 'G', 'NC-17', 'TV-Y7-FV', 'UR'], dtype=object)]
```

Q.9.1. How many Movies got The "TV-14" rating in Canada?

```
# columns : 'Movies', Ratings, County and using 'shape'
df[(df['Category']=='Movie') & (df['Rating']=='TV-14') & (df['Country']=='Canada')].shape

(11, 13)
```

Saved successfully!

Rating after YEar 2018?

```
df[(df['Category']=='TV Show') & (df['Rating']=='R') & (df['Year']>2018)]

#only 1 row
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	...
6437	s6436	TV Show	The Hateful Eight: Extended Version	Quentin Tarantino	Samuel L. Jackson, Kurt Russell, Jennifer L...	NaN	April 25, 2019	R	1 Season	TV Shows	Trapped at a stagecoach stop as a storm rages ...	

Q.10. What is the maximum duration of a Movie/Show on Netflix?

```
df['Duration'].nunique()

216

df['Duration'].dtypes

dtype('O')

#splitting [Duration ] >>> str.split()

df[['Minutes','Unit']]=df['Duration'].str.split(' ', expand= True)

df.head(2)
```

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano	Mexico	December 23, 2016	TV-MA	93 min	Dramas, International Movies	After a devastating earthquake hits Mexico Cit...

```
df[['Duration','Minutes','Unit']]
```

	Duration	Minutes	Unit
0	4 Seasons	4	Seasons
1	93 min	93	min
2	78 min	78	min
3	80 min	80	min
4	123 min	123	min
...	...	...	...
7784	99 min	99	min
7785	111 min	111	min
7786	44 min	44	min
7787	1 Season	1	Season
7788	90 min	90	min

7787 rows × 3 columns

```
# Changing Minutes Column into Numeric datatypes .
df['Minutes']=pd.to_numeric(df['Minutes'])

#Use MAX() to find maximum Duration.
df['Minutes'].max()
```

Saved successfully!

Q.11. Which Individual Country has the Highest No. of TV Shows?

```
#Creating a new DataFrame for TV Shows
df_tvshow=df[df['Category']=='TV Show']

df_tvshow.head(2)
```

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &... In a future where the elite inhabit an island ...
5	s6	TV Show	46	Serdar Akar	Erdal Beşikçioğlu, Yasemin Allen, Melis Birkan...	Turkey	July 1, 2017	TV-MA	1 Season	International TV Shows, TV Dramas, TV Mysteries A genetics professor experiments with a treatm...

```
from tables import index
df_tvshow.Country.value_counts().head(1)

#united states has Hughest numbers of Individual TV shows

United States    705
Name: Country, dtype: int64
```

Q.12. How can we sort the dataset by Year

```
df.sort_values(by='Year', ascending= False).head(4) #High>>>. Lower
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description
5564	s5564	Movie	Sherlock Holmes	Guy Ritchie	Robert Downey Jr., Jude Law, Rachel McAdams, M...	United States, Germany, United Kingdom, Australia	January 1, 2021	PG-13	128 min	Action & Adventure, Comedies	The ga afoot f ecoi detectiv
5919	s5919	Movie	Surf's Up	Ash Brannon, Chris Buck	Shia LaBeouf, Jeff Bridges, Zooey Deschanel, J...	United States	January 1, 2021	PG	86 min	Children & Family Movies, Comedies, Sports Movies	This O nomir anim comedy b
980	s981	TV Show	Bling Empire	NaN	NaN	NaN	January 15, 2021	TV-MA	1 Season	Reality TV	we Asian A
5955	s5955	Movie	Taare Zameen Par	Aamir Khan	Aamir Khan, Darsheel Safary, Tanay Chheda	India	January 10, 2021	PG	162 min	Dramas, International Movies	V daydre lsha so bo

Saved successfully! X