



Dataset Description::

This Netflix Dataset has information about the TV Shows and Movies available on Netflix.

It provides various metadata such as the type of content, cast, genres, country of origin, release details, and more. This dataset can be useful for content analysis, recommendation system development, or trend studies.

This dataset is collected from Flixable which is a third-party Netflix search engine.

```
In [1]: # Import the required libraries and dataset
```

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

```
In [3]: df=pd.read_csv("Netflix Dataset.csv")
```

```
In [4]: df.head() # top 5
```

Out[4]:

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

In [5]: `df.tail() # bottom 5`

Out[5]:

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

In [6]: `df.shape` # to show the No. of Rows and Columns

Out[6]: (7789, 11)

In [7]: # to show the No. of total values(elements) in the dataset
`df.size`

Out[7]: 85679

In [8]: # to show each column name
`df.columns`Out[8]: Index(['Show_Id', 'Category', 'Title', 'Director', 'Cast', 'Country',
'Release_Date', 'Rating', 'Duration', 'Type', 'Description'],
dtype='object')In [9]: # to show the data type of each columns
`df.dtypes`

```
Out[9]: 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
```

```
In [10]: df.info() # to show indexes, columns, data-types of each column, memory at once
```

```
<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
```

Check for duplicate records, if found then remove it

```
In [12]: df.head()
```

Out[12]:

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

In [13]: `df.shape`

Out[13]: (7789, 11)

In [14]: `df.duplicated()`

Out[14]:

```

0      False
1      False
2      False
3      False
4      False
...
7784   False
7785   False
7786   False
7787   False
7788   False
Length: 7789, dtype: bool

```

In [15]: `df[df.duplicated()] # to check row wise and detect the duplicate rows`

Out[15]:

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

In [16]: `df.drop_duplicates(inplace = True) # to remove the duplicate rows permanently`

In [17]: `df[df.duplicated()]`

Out[17]:

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	T
---------	----------	-------	----------	------	---------	--------------	--------	----------	---

In [18]: `df.shape`

Out[18]: (7787, 11)

Check for the Null Value

In [19]: `df.isnull() # to show where Null values is present`

Out[19]:

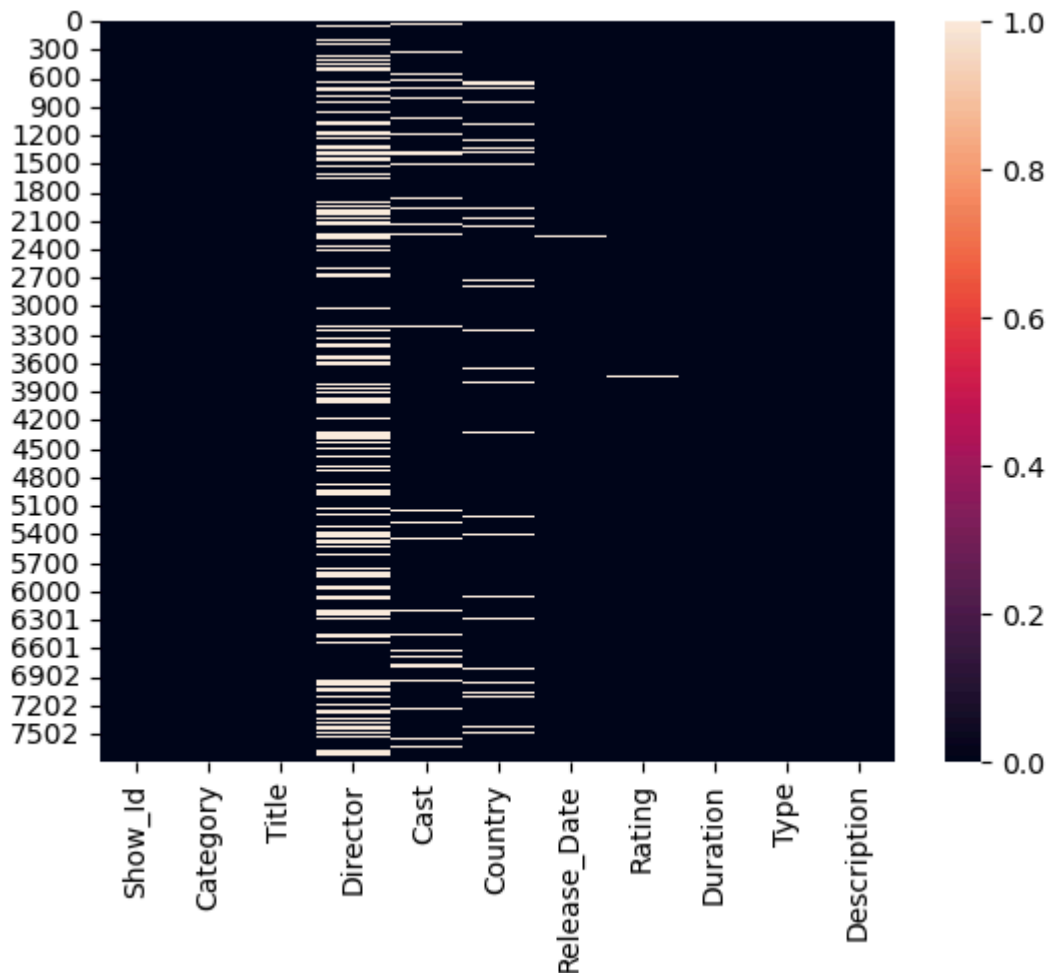
	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration
0	False	False	False	True	False	False	False	False	Fal
1	False	False	False	False	False	False	False	False	Fal
2	False	False	False	False	False	False	False	False	Fal
3	False	False	False	False	False	False	False	False	Fal
4	False	False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

7787 rows × 11 columns

In [20]: `df.isnull().sum()`

```
Out[20]: 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
```

In [21]: `# Graphical representation of null values``sns.heatmap(df.isnull())`Out[21]: `<Axes: >`



We got to know that the maximum number of Null values are present in Director column followed by Cast & Country

Now its time to get answer's for some questions

Q.1) For 'House of Cards', what is the Show Id and Who is the Director of this show ?

Q.2) In which year the highest number of the TV Shows & Movies were released ? Show with Bar Graph.

Q.3) How many Movies & TV Shows are in the dataset ? Show with Bar Graph.

Q.4) Show all the Movies that were released in year 2000.

Q.5) Show only the Titles of all TV Shows that were released in India only.

Q.6) Show Top 10 Directors, who gave the highest number of TV Shows & Movies to Netflix ?

Q.7) Show all the Records, where "Category is Movie and Type is Comedies" or "Country is United Kingdom".

Q.8) In how many movies/shows, Tom Cruise was cast ?

Q.9) What are the different Ratings defined by Netflix ?

- How many Movies got the 'TV-14' rating, in Canada ?
- How many TV Shows got the 'R' rating, after year 2018 ?

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

Q.11) Which individual country has the Highest No. of TV Shows ?

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

Q.13) Find all the instances where: Category is 'Movie' and Type is 'Dramas' or Category is 'TV Show' & Type is 'Kids' TV'.

In [22]: `df.head()`

Out[22]:

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

Q1. For 'House of Cards', what is the Show id and Who is the director of this show?

In [24]: `df[df["Title"].isin(["House of Cards"])] # to show all records of a particular i`

Out[24]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Di
2832	s2833	TV Show	House of Cards	Robin Wright, David Fincher, Gerald McRaney, J...	Kevin Spacey, Robin Wright, Kate Mara, Corey S...	United States	November 2, 2018	TV-MA	S

In [25]: `df[df['Title'].str.contains('House of Cards')] # to show all records of a partic`

Out[25]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Di
2832	s2833	TV Show	House of Cards	Robin Wright, David Fincher, Gerald McRaney, J...	Kevin Spacey, Robin Wright, Kate Mara, Corey S...	United States	November 2, 2018	TV-MA	S

Q2. In which year highest number of the TV Shows & Movies were released? Show with bar graph.

In [26]: `df.dtypes`

```
Out[26]: 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
```

```
In [27]: # Create new column - df["Date_N"]
# Pass the original column "Release Date" in which we have to work

df["Date_N"] = pd.to_datetime(df["Release_Date"], errors="coerce")
```

In [28]: `df["Date_N"]`

```
Out[28]: 0      2020-08-14
         1      2016-12-23
         2      2018-12-20
         3      2017-11-16
         4      2020-01-01
         ...
        7784    2020-10-19
        7785    2019-03-02
        7786    2020-09-25
        7787    2020-10-31
        7788    2020-03-01
        Name: Date_N, Length: 7787, dtype: datetime64[ns]
```

```
In [29]: df.dtypes
```

```
Out[29]: 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
```

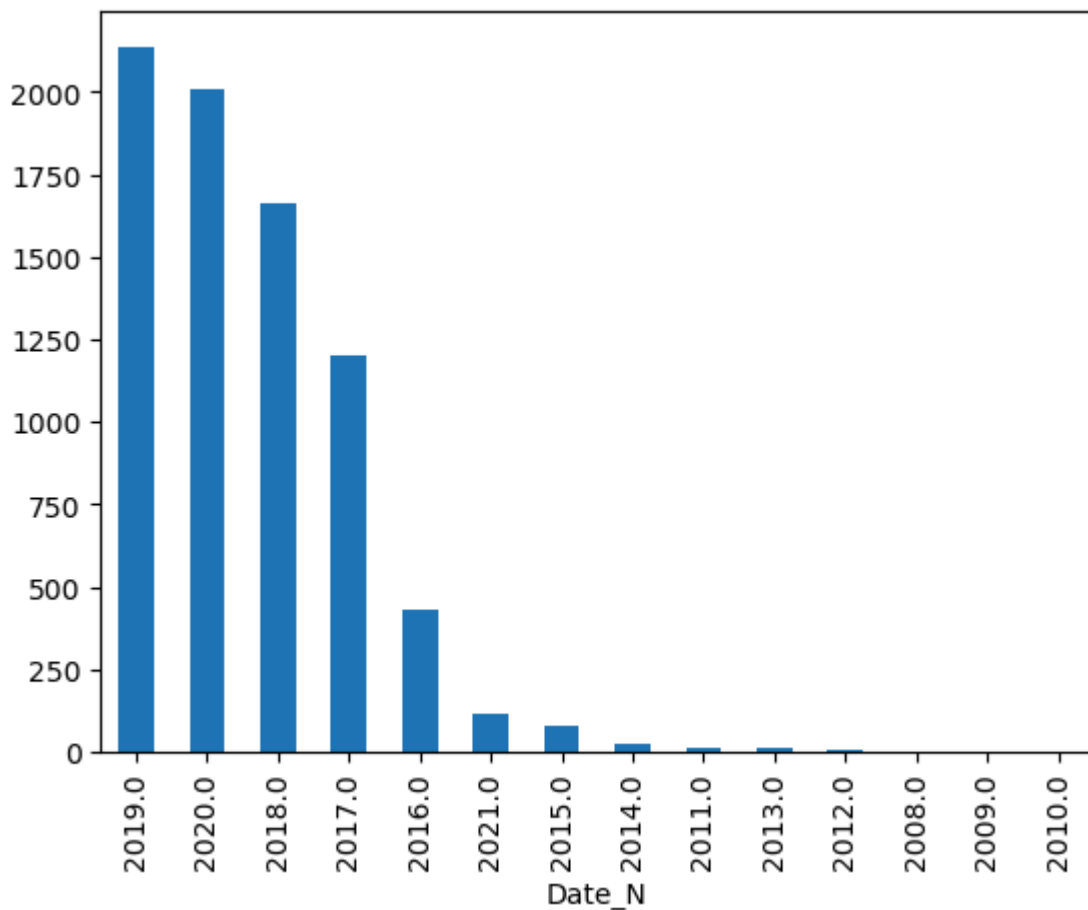
```
In [30]: df['Date_N'].dt.year.value_counts()
```

```
Out[30]: Date_N
         2019.0    2136
         2020.0    2008
         2018.0    1661
         2017.0    1201
         2016.0     432
         2021.0     117
         2015.0      79
         2014.0      24
         2011.0      13
         2013.0      10
         2012.0       3
         2008.0       2
         2009.0       2
         2010.0       1
         Name: count, dtype: int64
```

```
In [31]: # Graphical Representation using Bar Graph
```

```
df['Date_N'].dt.year.value_counts().plot(kind='bar') # It counts the occurrence
```

```
Out[31]: <Axes: xlabel='Date_N'>
```



Q 3. How many Movies & TV Shows are in the dataset? Show with bar graph.

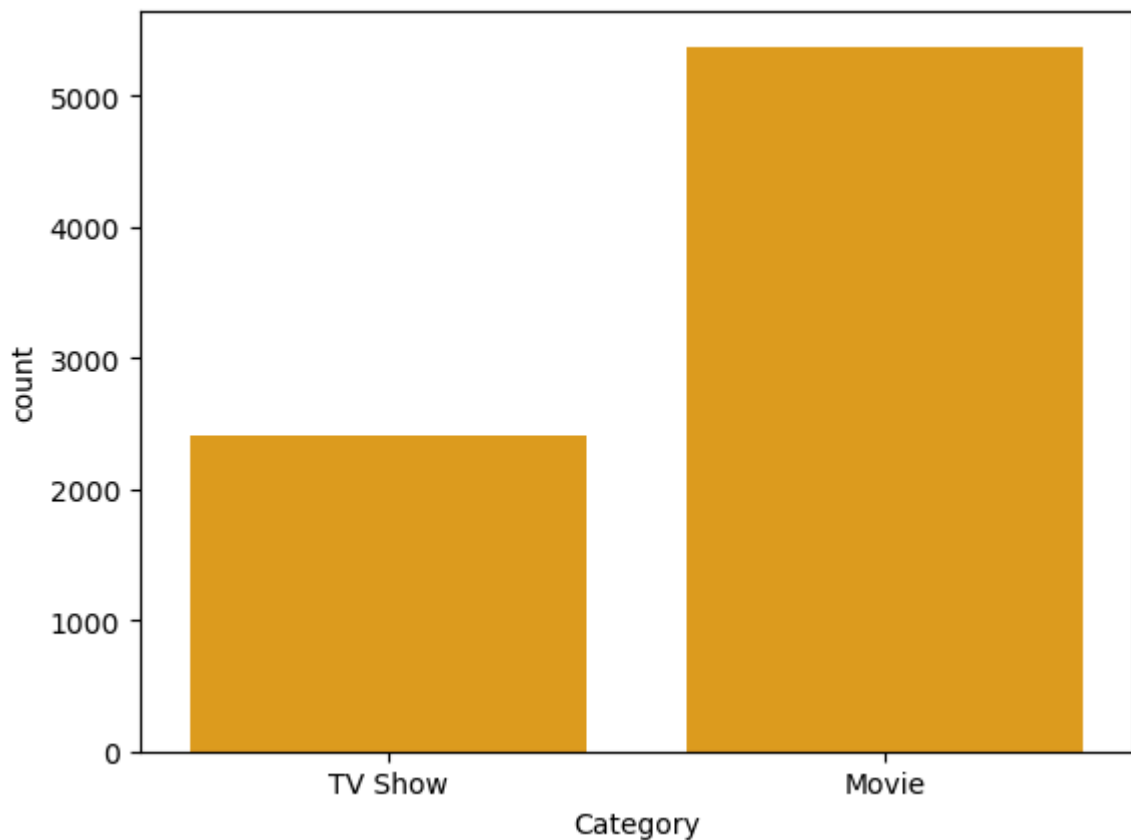
```
In [33]: df.groupby('Category').Category.count()
```

```
Out[33]: Category
Movie      5377
TV Show    2410
Name: Category, dtype: int64
```

- To group all unique items of a column and show their count
- Use groupby for all movies & tv shows from "category" column -
data.groupby('Category')
- Total count for all movies and tv shows - .Category.count()

```
In [36]: # to show the count of all unique values of any column in the form of bar graph
sns.countplot(x="Category", data=df, color="orange")
```

```
Out[36]: <Axes: xlabel='Category', ylabel='count'>
```



Q4. Show all the Movies that were released in year 2020 ?

In [38]: `df['Year'] = df['Date_N'].dt.year # to create new Year column, it will consider`

In [39]: `df.head(1)`

Out[39]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Dur:
--	---------	----------	-------	----------	------	---------	--------------	--------	------

0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	See
---	----	---------	----	-----	---	--------	-----------------	-------	-----



In [41]: `df[(df["Category"]=="Movie") & (df["Year"]==2020)]`

Out[41]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Ra
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	Pe
6	s7	Movie	122	Yasir Al Yasiri	Amina Khalil, Ahmed Dawood, Tarek Lotfy, Ahmed...	Egypt	June 1, 2020	
14	s15	Movie	3022	John Suits	Omar Epps, Kate Walsh, Miranda Cosgrove, Angus...	United States	March 19, 2020	
27	s28	Movie	#Alive	Cho Il	Yoo Ah-in, Park Shin-hye	South Korea	September 8, 2020	
28	s29	Movie	#AnneFrank - Parallel Stories	Sabina Fedeli, Anna Migotto	Helen Mirren, Gengher Gatti	Italy	July 1, 2020	T
...
7762	s7761	Movie	Zaki Chan	Wael Ihsan	Ahmed Helmy, Yasmin Abdulaziz, Hassan Hosny, H...	Egypt	May 19, 2020	TV
7783	s7782	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	
7784	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car...	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	
7786	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Ra
7788	s7787	Movie	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	Sam Dunn	NaN	United Kingdom, Canada, United States	March 1, 2020

1312 rows × 13 columns

Q 5. Show only the Titles of all TV Shows that were released in India only.

```
In [42]: df[(df['Category']=='TV Show') & (df['Country']=='India')]['Title']
```

```
Out[42]: 86          21 Sarfarosh: Saragarhi 1897
132          7 (Seven)
340          Agent Raghav
364          Akbar Birbal
533          Anjaan: Rural Myths
...
6249          The Creative Indians
6400  The Golden Years with Javed Akhtar
6469          The House That Made Me
7294          Typewriter
7705          Yeh Meri Family
Name: Title, Length: 71, dtype: object
```

Q 6. Show top 10 directors, who gave the highest number of TV Shows & Movies to Netflix ?

```
In [44]: df['Director'].value_counts().head(10)
```

```
Out[44]: Director
Raúl Campos, Jan Suter    18
Marcus Raboy              16
Jay Karas                 14
Cathy Garcia-Molina      13
Youssef Chahine           12
Jay Chapman               12
Martin Scorsese           12
Steven Spielberg          10
David Dhawan              9
Lance Bangs               8
Name: count, dtype: int64
```

Q7: Show all the Records, where "Category is Movie and Type is Comedies" or "Country is United Kingdom" ?

```
In [46]: df[(df['Category']=='Movie') & (df['Type']=='Comedies')]
```

Out[46]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Re
33	s34	Movie	#realityhigh	Fernando Lebrija	Nesta Cooper, Kate Walsh, John Michael Higgins...	United States	September 8, 2017	T
72	s73	Movie	17 Again	Burr Steers	Zac Efron, Leslie Mann, Matthew Perry, Thomas ...	United States	January 1, 2021	P
84	s85	Movie	21 & Over	Jon Lucas, Scott Moore	Miles Teller, Skylar Astin, Justin Chon, Sarah...	United States	April 16, 2019	
148	s149	Movie	A Bad Moms Christmas	Jon Lucas, Scott Moore	Mila Kunis, Kristen Bell, Kathryn Hahn, Susan ...	United States, China	February 14, 2020	
174	s175	Movie	A Futile and Stupid Gesture	David Wain	Will Forte, Domhnall Gleeson, Martin Mull, Joe...	United States	January 26, 2018	
...
7341	s7340	Movie	Undercover Brother 2	Leslie Small	Michael Jai White, Vince Swann, Barry Bostwick...	United States	November 5, 2019	
7590	s7589	Movie	Where's the Money	Scott Zabielski	Andrew Bachelor, Kat Graham, Mike Epps, Terry ...	United States	January 22, 2018	

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rea	
7623	s7622	Movie	Wild Oats	Andy Tennant	Shirley MacLaine, Jessica Lange, Billy Connoll...	Germany, United States	May 13, 2017	P
7634	s7633	Movie	Wine Country	Amy Poehler	Amy Poehler, Maya Rudolph, Ana Gasteyer, Rache...	United States	May 10, 2019	
7664	s7663	Movie	Work It	Laura Terruso	Sabrina Carpenter, Jordan Fisher, Liza Koshy, ...	United States	August 7, 2020	T

97 rows × 13 columns

In [47]: `df[(df['Category']=='Movie') & (df['Type']=='Comedies') | (df['Country']=='Unite`

Out[47]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	I
19	s20	Movie	'89	NaN	Lee Dixon, Ian Wright, Paul Merson	United Kingdom	May 16, 2018	
33	s34	Movie	#realityhigh	Fernando Lebrija	Nesta Cooper, Kate Walsh, John Michael Higgins...	United States	September 8, 2017	
58	s59	TV Show	100% Hotter	NaN	Daniel Palmer, Melissa Sophia, Karen Williams,...	United Kingdom	November 1, 2019	
72	s73	Movie	17 Again	Burr Steers	Zac Efron, Leslie Mann, Matthew Perry, Thomas ...	United States	January 1, 2021	
82	s83	Movie	2036 Origin Unknown	Hasraf Dulull	Katee Sackhoff, Ray Fearon, Julie Cox, Steven ...	United Kingdom	December 20, 2018	
...
7670	s7669	TV Show	World War II in Colour	NaN	Robert Powell	United Kingdom	August 1, 2017	
7671	s7670	TV Show	World's Busiest Cities	NaN	Anita Rani, Ade Adepitan, Dan Snow	United Kingdom	February 1, 2019	
7688	s7687	Movie	XV: Beyond the Tryline	Pierre Deschamps	NaN	United Kingdom	March 18, 2020	
7725	s7724	Movie	You Can Tutu	James Brown	Lily O'Regan, Jeannettsy Enriquez	United Kingdom	December 31, 2017	

Show_Id	Category	Title	Director	Cast	Country	Release_Date	I
				Borges, Joel...			
7740	s7739	TV Show	Young Wallander	NaN	Adam Pålsson, Richard Dillane, Leanne Best, El...	United Kingdom	September 3, 2020

485 rows × 13 columns

Q 8. In how many movies/shows, Tom Cruise was cast?

```
In [48]: df[df['Cast']=='Tom Cruise']
```

```
Out[48]:
```

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	T
---------	----------	-------	----------	------	---------	--------------	--------	----------	---

```
In [49]: df[df['Cast'].str.contains('Tom Cruise')]
```

```
-----
ValueError                                Traceback (most recent call last)
Cell In[49], line 1
----> 1 df[df['Cast'].str.contains('Tom Cruise')]

File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:4092, in DataFrame.__getitem__(self, key)
    4089     return self.where(key)
    4091 # Do we have a (boolean) 1d indexer?
-> 4092 if com.is_bool_indexer(key):
    4093     return self._getitem_bool_array(key)
    4095 # We are left with two options: a single key, and a collection of keys,
    4096 # We interpret tuples as collections only for non-MultiIndex

File ~\anaconda3\Lib\site-packages\pandas\core\common.py:136, in is_bool_indexer(key)
    132     na_msg = "Cannot mask with non-boolean array containing NA / NaN values"
    133     if lib.is_bool_array(key_array, skipna=True):
    134         # Don't raise on e.g. ["A", "B", np.nan], see
    135         # test_loc_getitem_list_of_labels_categoricalindex_with_na
--> 136         raise ValueError(na_msg)
    137     return False
    138 return True

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

Creating new data-frame

```
In [50]: # it drop the rows that contains all missing values

df_new = df.dropna()
```

In [51]: `df_new.head(2)`

Out[51]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Dura
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78

In [52]: `df_new[df_new['Cast'].str.contains('Tom Cruise')]`

Out[52]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating
3860	s3861	Movie	Magnolia	Paul Thomas Anderson	Jeremy Blackman, Tom Cruise, Melinda Dillon, A...	United States	January 1, 2020	
5071	s5071	Movie	Rain Man	Barry Levinson	Dustin Hoffman, Tom Cruise, Valeria Golino, Ge...	United States	July 1, 2019	

Q. 9 What are the different Ratings defined by Netflix?

In [53]: `df["Rating"].nunique()`

Out[53]: 14

In [54]: `df["Rating"].unique()`

Out[54]: 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)

In [55]: `### 9.1 How many Movies got the 'TV-14' rating in Canada ?`

```
df[(df['Category']=='Movie') & (df['Rating']=='TV-14')].shape
```

Out[55]: (1272, 13)

In [56]: `df[(df['Category']=='Movie') & (df['Rating']=='TV-14') & (df['Country']=='Canada`

Out[56]: (11, 13)

In [57]: `### Q 9.2 How many TV Show got the 'R' rating, after year 2018 ?`

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

Out[57]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating
--	---------	----------	-------	----------	------	---------	--------------	--------

4497	s4498	TV Show	Nobel	NaN	Aksel Hennie, Tuva Novotny, Anders Danielsen L...	Norway	December 10, 2016	R
------	-------	---------	-------	-----	---	--------	-------------------	---

6437	s6436	TV Show	The Hateful Eight: Extended Version	Quentin Tarantino	Samuel L. Jackson, Kurt Russell, Jennifer Jaso...	NaN	April 25, 2019	R
------	-------	---------	-------------------------------------	-------------------	---	-----	----------------	---



In [59]: `df[(df['Category']=='TV Show') & (df['Rating']=='R') & (df['Year']>2018)]`

Out[59]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating
--	---------	----------	-------	----------	------	---------	--------------	--------

6437	s6436	TV Show	The Hateful Eight: Extended Version	Quentin Tarantino	Samuel L. Jackson, Kurt Russell, Jennifer Jaso...	NaN	April 25, 2019	R
------	-------	---------	-------------------------------------	-------------------	---	-----	----------------	---



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

In [60]: `df['Duration'].unique()`

```
Out[60]: array(['4 Seasons', '93 min', '78 min', '80 min', '123 min', '1 Season',
              '95 min', '119 min', '118 min', '143 min', '103 min', '89 min',
              '91 min', '149 min', '144 min', '124 min', '87 min', '110 min',
              '128 min', '117 min', '100 min', '2 Seasons', '84 min', '99 min',
              '90 min', '102 min', '104 min', '105 min', '56 min', '125 min',
              '81 min', '97 min', '106 min', '107 min', '109 min', '44 min',
              '75 min', '101 min', '3 Seasons', '37 min', '113 min', '114 min',
              '130 min', '94 min', '140 min', '135 min', '82 min', '70 min',
              '121 min', '92 min', '164 min', '53 min', '83 min', '116 min',
              '86 min', '120 min', '96 min', '126 min', '129 min', '77 min',
              '137 min', '148 min', '28 min', '122 min', '176 min', '85 min',
              '22 min', '68 min', '111 min', '29 min', '142 min', '168 min',
              '21 min', '59 min', '20 min', '98 min', '108 min', '76 min',
              '26 min', '156 min', '30 min', '57 min', '150 min', '133 min',
              '115 min', '154 min', '127 min', '146 min', '136 min', '88 min',
              '131 min', '24 min', '112 min', '74 min', '63 min', '38 min',
              '25 min', '174 min', '60 min', '153 min', '158 min', '151 min',
              '162 min', '54 min', '51 min', '69 min', '64 min', '147 min',
              '42 min', '79 min', '5 Seasons', '40 min', '45 min', '172 min',
              '10 min', '163 min', '9 Seasons', '55 min', '72 min', '61 min',
              '71 min', '160 min', '171 min', '48 min', '139 min', '157 min',
              '15 min', '65 min', '134 min', '161 min', '62 min', '8 Seasons',
              '186 min', '49 min', '73 min', '58 min', '165 min', '166 min',
              '138 min', '159 min', '141 min', '132 min', '52 min', '67 min',
              '34 min', '66 min', '312 min', '180 min', '47 min', '6 Seasons',
              '155 min', '14 min', '177 min', '11 min', '9 min', '46 min',
              '145 min', '11 Seasons', '7 Seasons', '13 Seasons', '8 min',
              '12 min', '12 Seasons', '10 Seasons', '43 min', '50 min', '23 min',
              '185 min', '200 min', '169 min', '27 min', '170 min', '196 min',
              '33 min', '181 min', '204 min', '32 min', '35 min', '167 min',
              '16 Seasons', '179 min', '193 min', '13 min', '214 min', '17 min',
              '173 min', '192 min', '209 min', '187 min', '41 min', '182 min',
              '224 min', '233 min', '189 min', '152 min', '19 min', '15 Seasons',
              '208 min', '237 min', '31 min', '178 min', '230 min', '194 min',
              '228 min', '195 min', '3 min', '16 min', '5 min', '18 min',
              '205 min', '190 min', '36 min', '201 min', '253 min', '203 min',
              '191 min'], dtype=object)
```

```
In [61]: df['Duration'].unique()
```

```
Out[61]: 216
```

```
In [62]: df.Duration.dtypes
```

```
Out[62]: dtype('O')
```

```
In [63]: df[['Minutes', 'Unit']] = df['Duration'].str.split(' ', expand=True) # # str.spl
```

```
In [64]: df.Minutes.max()
```

```
Out[64]: '99'
```

```
In [65]: df.Minutes.min()
```

```
Out[65]: '1'
```

Q 11 Which country has the highest number of TV Shows?

```
In [66]: # We have to consider "Country" and "Category" columns.
# Need to create a new data frame which only contain TV Shows.
df_tvshow = df[df['Category']=='TV Show']
```

```
In [67]: df_tvshow.head(2)
```

```
Out[67]:
```

	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	Se...
5	s6	TV Show	46	Serdar Akar	Erdal Beşikçioğlu, Yasemin Allen, Melis Birkan...	Turkey	July 1, 2017	TV-MA	1 Se...

```
In [68]: df_tvshow.Country.value_counts()
```

```
Out[68]: Country
United States          705
United Kingdom         204
Japan                  157
South Korea            147
India                   71
...
United Kingdom, Ireland, United States    1
United Kingdom, Spain                     1
Italy, United States                       1
Japan, Canada                             1
United States, France, South Korea, Indonesia  1
Name: count, Length: 183, dtype: int64
```

- United States has the highest number of TV shows followed by United Kingdom and Japan

Q 12 How can we sort the dataset by Year?

```
In [69]: df.sort_values(by="Year").head()
```

Out[69]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rat
7115	s7114	Movie	To and From New York	Sorin Dan Mihalcescu	Barbara King, Shaana Diya, John Krisiukenas, Y...	United States	January 1, 2008	
1765	s1766	TV Show	Dinner for Five	NaN	NaN	United States	February 4, 2008	
5766	s5766	Movie	Splatter	Joe Dante	Corey Feldman, Tony Todd, Tara Leigh, Erin Way...	United States	November 18, 2009	
3248	s3249	Movie	Just Another Love Story	Ole Bornedal	Anders W. Berthelsen, Rebecka Hemse, Nikolaj L...	Denmark	May 5, 2009	
3840	s3841	Movie	Mad Ron's Prevues from Hell	Jim Monaco	Nick Pawlow, Jordu Schell, Jay Kushwara, Micha...	United States	November 1, 2010	

In [70]: `df.sort_values(by="Year", ascending = False).head()`

Out[70]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Ratin
601	s602	Movie	Asphalt Burning	Hallvard Bræin	Anders Baasmo, Kathrine Thorborg Johansen, Ida...	Norway	January 2, 2021	TV-1
5060	s5060	Movie	Radium Girls	Lydia Dean Pilcher, Ginny Mohler	Joey King, Abby Quinn, Cara Seymour, Scott She...	United States	January 16, 2021	TV-1
980	s981	TV Show	Bling Empire	NaN	NaN	NaN	January 15, 2021	TV-M
3237	s3238	Movie	Julie & Julia	Nora Ephron	Meryl Streep, Amy Adams, Stanley Tucci, Chris ...	United States	January 1, 2021	PG-1
5305	s5305	TV Show	Running Man	NaN	Kim Seo-yeong, Um Sang-hyun, Kwon Changwook, K...	NaN	January 1, 2021	TV-Y

Q 13) Find all instances where:

- Category is 'Movie' and Type is 'Dramas'
- Category is 'TV Show' and Type is 'Kids TV'

In [72]: `df[(df['Category']=='Movie') & (df['Type']=='Dramas')].head(2)`

Out[72]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Durat
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123
7	s8	Movie	187	Kevin Reynolds	Samuel L. Jackson, John Heard, Kelly Rowan, Cl...	United States	November 1, 2019	R	119

In [74]: `df[(df['Category']=='Movie') & (df['Type']=='Dramas')] | (df['Category']=='TV Sho`

Out[74]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Durat
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123
7	s8	Movie	187	Kevin Reynolds	Samuel L. Jackson, John Heard, Kelly Rowan, Cl...	United States	November 1, 2019	R	119

Commands used in this project

- `head()` - It shows the first N rows in the data (by default, N=5).
- `tail()` - It shows the last N rows in the data (by default, N=5).
- `shape` - It shows the total no. of rows and no. of columns of the dataframe.
- `size` - To show No. of total values(elements) in the dataset.
- `columns` - To show each Column Name.
- `dtypes` - To show the data-type of each column.
- `info()` - To show indexes, columns, data-types of each column, memory at once.
- `value_counts` - In a column, it shows all the unique values with their count. It can be applied on a single column only.
- `unique()` - It shows the all unique values of the series.
- `nunique()` - It shows the total no. of unique values in the series.
- `duplicated()` - To check row wise and detect the Duplicate rows.

- `isnull()` - To show where Null value is present.
- `dropna()` - It drops the rows that contains all missing values.
- `isin()` - To show all records including particular elements.
- `str.contains()` - To get all records that contains a given string.
- `str.split()` - It splits a column's string into different columns.
- `to_datetime()` - Converts the data-type of Date-Time Column into `datetime[ns]` datatype.
- `dt.year.value_counts()` - It counts the occurrence of all individual years in Time column.
- `groupby()` - Groupby is used to split the data into groups based on some criteria.
- `sns.countplot(df['Col_name'])` - To show the count of all unique values of any column in the form of bar graph.
- `max()`, `min()` - It shows the maximum/minimum value of the series.
- `mean()` - It shows the mean value of the series.