

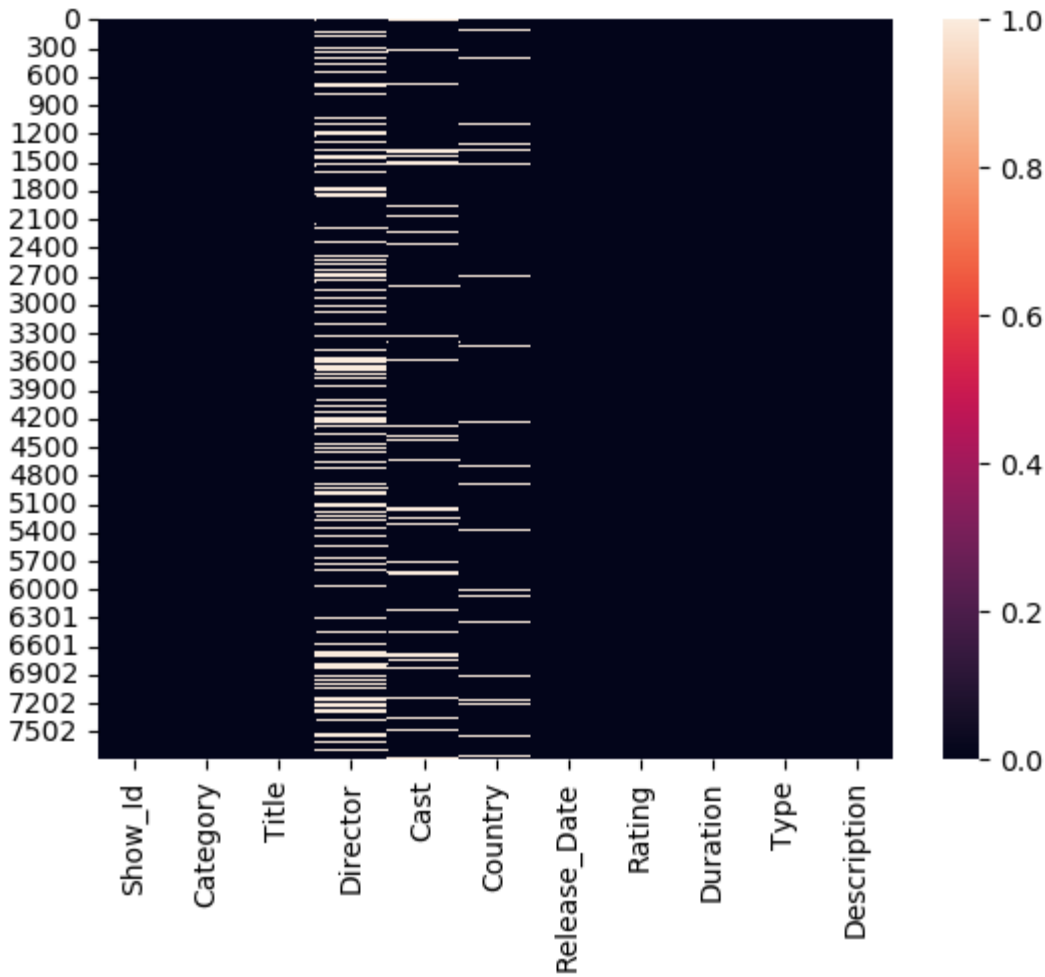
```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [2]: df = pd.read_csv("./netflix.csv")
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
```

```
In [3]: #Task. 1) Is there any Duplicate Record in this dataset ? If yes, then remove
df.duplicated().sum()
df.drop_duplicates(inplace = True)
```

```
In [4]: #Task. 2) Is there any Null Value present in any column ? Show with Heat-map.
sns.heatmap(df.isnull())
plt.show()
```



In [5]:

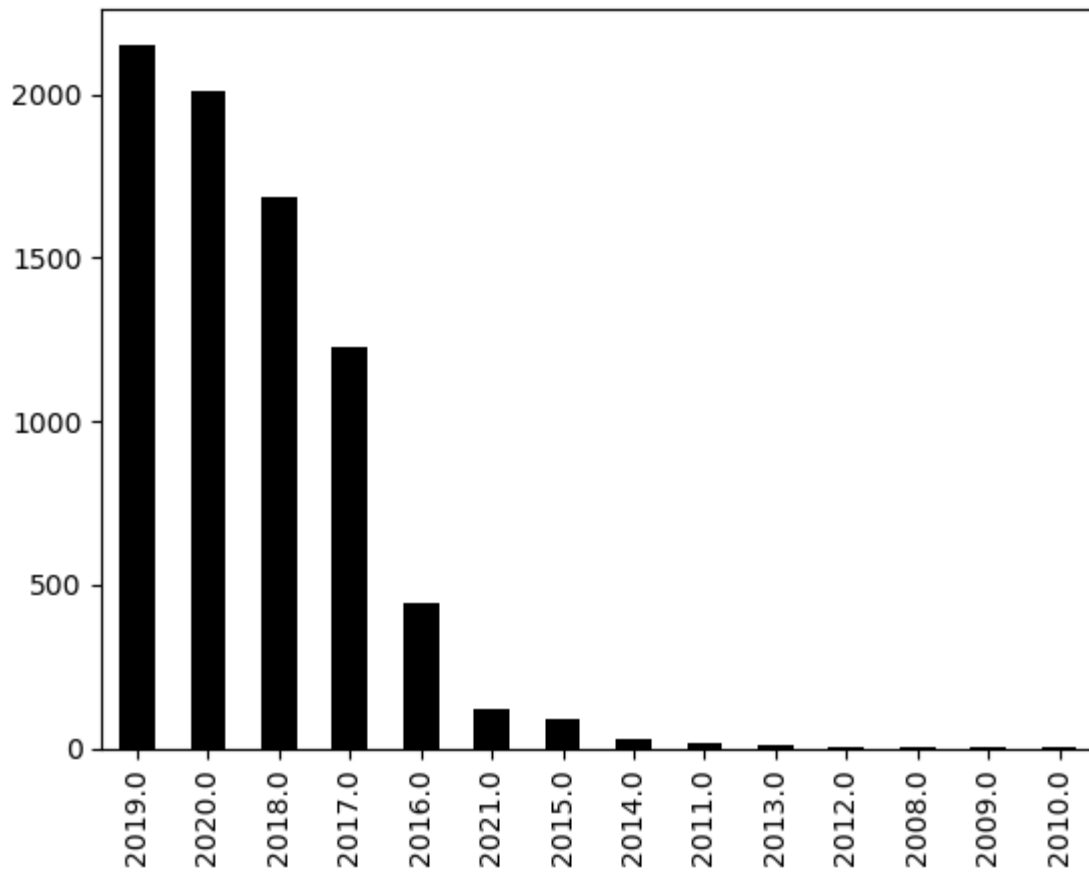
```
#Q. 1) For 'House of Cards', what is the Show Id and Who is the Director of t
df[df["Title"] == "House of Cards"]
#df[df["Title"].str.contains("House of Cards")]
#df[df["Title"].isin(["House of Cards"])]
```

Out[5]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	
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	6 Seasons	Dra Thr

In [6]:

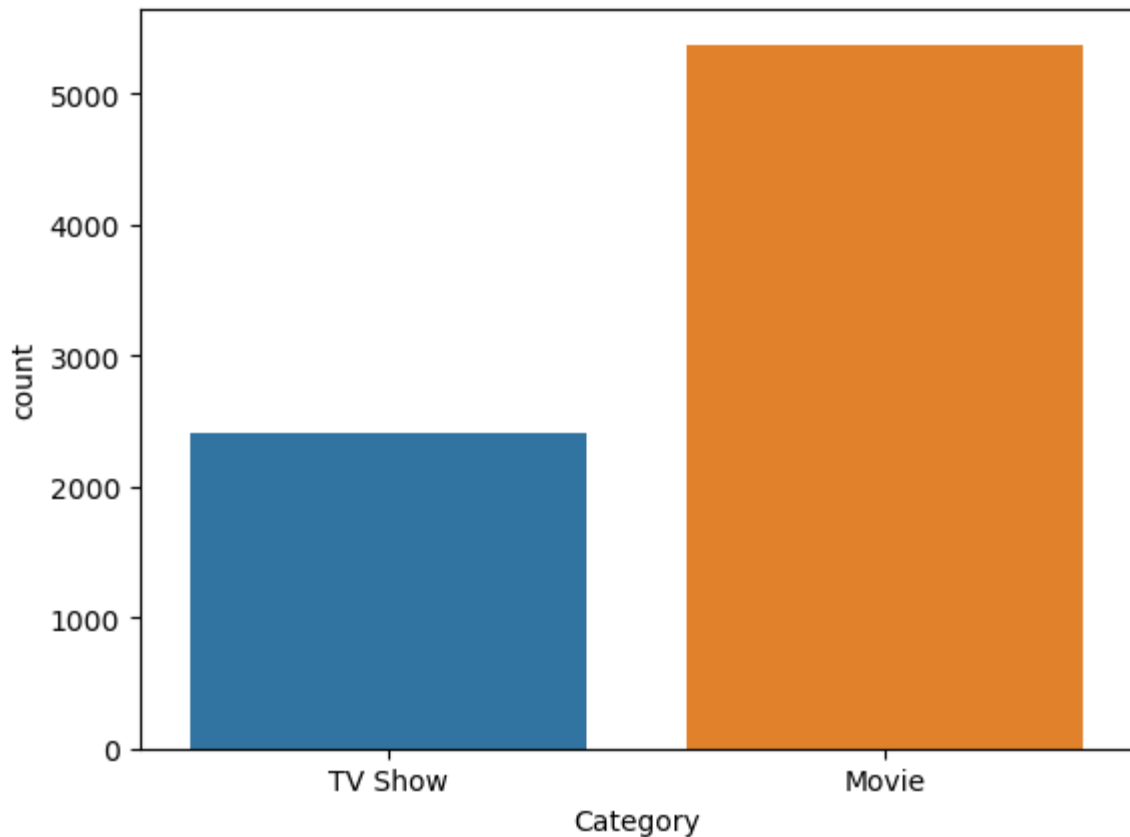
```
#Q. 2) In which year the highest number of the TV Shows & Movies were release
df["Release_Date"] = pd.to_datetime(df.Release_Date)
df.Release_Date.dt.year.value_counts().plot(kind = "bar", color = "black")
plt.show()
```



```
In [7]: #Q. 3) How many Movies & TV Shows are in the dataset ? Show with Bar Graph.  
df.Category.value_counts()
```

```
Out[7]: Movie      5377  
TV Show    2410  
Name: Category, dtype: int64
```

```
In [8]: sns.countplot(x = df.Category)  
plt.show()
```



In [9]:

```
#Q. 4) Show all the Movies that were released in year 2000.
df[(df.Release_Date.dt.year == 2000) & (df.Category == "Movie")]
```

Out[9]:

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

In [10]:

```
#Q. 5) Show only the Titles of all TV Shows that were released in Germany only
df[(df.Country == "Germany") & (df["Category"] == "TV Show")]["Title"]
```

Out[10]:

213	A Perfect Crime
672	Babylon Berlin
731	Barbarians
931	Biohackers
1277	Charité
1539	Criminal: Germany
1797	Dogs of Berlin
2261	Frank Elstner: Just One Last Question
2774	Holiday Secrets
2849	How to Sell Drugs Online (Fast)
4388	Nailed It! Germany
4546	NSU German History X
4711	Over Christmas
4829	Perfume
5623	Sing On! Germany
5659	Skylines
6575	The Last Word
7357	Unorthodox
7502	We Are the Wave

Name: Title, dtype: object

```
In [11]: #Q. 6) Show Top 10 Directors, who gave the highest number of TV Shows & Movie
df["Director"].value_counts(ascending = False)[0:10]
#df["Director"].value_counts(ascending = False).head(10)
```

```
Out[11]: Raúl Campos, Jan Suter      18
Marcus Raboy                      16
Jay Karas                         14
Cathy Garcia-Molina              13
Jay Chapman                      12
Youssef Chahine                  12
Martin Scorsese                  12
Steven Spielberg                 10
David Dhawan                     9
Hakan Algül                      8
Name: Director, dtype: int64
```

```
In [12]: #Q. 7) Show all the Records, where "Category is Movie and Type is Comedies" o
df[(df["Category"] == "Movie") & (df["Type"] == "Comedies") | (df["Country"]
```

```
Out[12]:
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Dura
19	s20	Movie	'89	NaN	Lee Dixon, Ian Wright, Paul Merson	United Kingdom	2018-05-16	TV-PG	87
33	s34	Movie	#realityhigh	Fernando Lebrija	Nesta Cooper, Kate Walsh, John Michael Higgins...	United States	2017-09-08	TV-14	99
58	s59	TV Show	100% Hotter	NaN	Daniel Palmer, Melissa Sophia, Karen Williams,...	United Kingdom	2019-11-01	TV-14	See
72	s73	Movie	17 Again	Burr Steers	Zac Efron, Leslie Mann, Matthew Perry, Thomas ...	United States	2021-01-01	PG-13	102
82	s83	Movie	2036 Origin Unknown	Hasraf Dulull	Katee Sackhoff, Ray Fearon, Julie Cox, Steven ...	United Kingdom	2018-12-20	TV-14	95
...	...	...	...	...	...	...	...	...	...
7670	s7669	TV Show	World War II in Colour	NaN	Robert Powell	United Kingdom	2017-08-01	TV-MA	See

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration
7671	s7670	TV Show	World's Busiest Cities	NaN	Anita Rani, Ade Adepitan, Dan Snow	United Kingdom	2019-02-01	TV-PG	See
7688	s7687	Movie	XV: Beyond the Tryline	Pierre Deschamps	NaN	United Kingdom	2020-03-18	TV-14	91
7725	s7724	Movie	You Can Tutu	James Brown	Lily O'Regan, Jeannettsy Enriquez Borges, Joel...	United Kingdom	2017-12-31	TV-G	87
7740	s7739	TV Show	Young Wallander	NaN	Adam Pålsson, Richard Dillane, Leanne Best, El...	United Kingdom	2020-09-03	TV-MA	See

```
In [13]: #Q. 8) In how many movies/shows, Tom Cruise was cast ?
df_drop = df.dropna()
df_drop[df_drop["Cast"].str.contains("Tom Cruise")]
```

```
Out[13]:
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration
3860	s3861	Movie	Magnolia	Paul Thomas Anderson	Jeremy Blackman, Tom Cruise, Melinda Dillon, A...	United States	2020-01-01	R	189 min
5071	s5071	Movie	Rain Man	Barry Levinson	Dustin Hoffman, Tom Cruise, Valeria Golino, Ge...	United States	2019-07-01	R	134 min

```
In [14]: #Q. 9) What are the different Ratings defined by Netflix ?

#df.Rating.unique()
df.Rating.value_counts()
```

```
Out[14]: TV-MA      2863
TV-14      1931
TV-PG      806
R          665
```

PG-13	386
TV-Y	280
TV-Y7	271
PG	247
TV-G	194
NR	84
G	39
TV-Y7-FV	6
UR	5
NC-17	3

```
In [15]: ##Q. 9.1) How many Movies got the 'TV-14' rating, in Canada ?
#df[(df.Rating == "TV-14") & (df.Category == "Movie") & (df.Country == "Canada")]
len(df[(df.Rating == "TV-14") & (df.Category == "Movie") & (df.Country == "Canada")])
#df[(df.Rating == "TV-14") & (df.Category == "Movie") & (df.Country == "Canada")]
#df[(df.Rating == "TV-14") & (df.Category == "Movie") & (df.Country == "Canada")]
```

Out[15]: 11

```
In [16]: ##Q. 9.2) How many TV Shows got the 'R' rating, after year 2018 ?
#df[(df.Release_Date.dt.year >= 2018) & (df.Rating == "R") & (df.Category == "TV Show")]
#len(df[(df.Release_Date.dt.year >= 2018) & (df.Rating == "R") & (df.Category == "TV Show")])
df[(df.Release_Date.dt.year >= 2018) & (df.Rating == "R") & (df.Category == "TV Show")]
```

Out[16]: (11,)

```
In [17]: #Q. 10) What is the maximum duration of a Movie/Show on Netflix ?
df[["minutes", "unit"]] = df["Duration"].str.split(" ", expand = True)
df.minutes.astype(int).groupby(df.Category).max()
```

```
Out[17]: Category
Movie      312
TV Show     16
Name: minutes, dtype: int64
```

```
In [18]: #Q. 11) Which individual country has the Highest No. of TV Shows ?
df[df["Category"] == "TV Show"]["Country"].value_counts()
```

```
Out[18]: United States      705
United Kingdom      204
Japan      157
South Korea      147
India      71
...
Canada, United States, United Kingdom, France, Luxembourg      1
United States, Italy      1
Chile, Italy      1
Canada, United Kingdom      1
United States, France, South Korea, Indonesia      1
Name: Country, Length: 183, dtype: int64
```

```
In [19]: #Q. 12) How can we sort the dataset by Year ?
df.sort_values(by = "Release_Date", ascending = False)
```

Out[19]:	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Dural
1673	s1674	Movie	Death of Me	Darren Lynn Bousman	Maggie Q, Luke Hemsworth, Alex Essoe, Ingkarat...	United States, Thailand	2021-01-16	R	94
5060	s5060	Movie	Radium Girls	Lydia Dean Pilcher, Ginny Mohler	Joey King, Abby Quinn, Cara Seymour, Scott She...	United States	2021-01-16	TV-14	103
205	s206	Movie	A Monster Calls	J.A. Bayona	Lewis MacDougall, Sigourney Weaver, Felicity J...	United Kingdom, Spain, United States	2021-01-16	PG-13	108
1832	s1833	Movie	Double Dad	Cris D'Amato	Maisa Silva, Eduardo Moscovis, Marcelo Médici,...	Brazil	2021-01-15	TV-G	105
4710	s4711	Movie	Outside the Wire	Mikael Håfström	Anthony Mackie, Damson Idris, Emily Beecham, M...	NaN	2021-01-15	R	116
...	...	...	...	...	...	...	...	...	
3374	s3375	TV Show	Kikoriki	NaN	Igor Dmitriev	NaN	NaT	TV-Y	Seas
3492	s3493	TV Show	La Familia P. Luche	NaN	Eugenio Derbez, Consuelo Duval, Luis Manuel Áv...	United States	NaT	TV-14	Seas
3946	s3947	TV Show	Maron	NaN	Marc Maron, Judd Hirsch, Josh Brener, Nora Zeh...	United States	NaT	TV-MA	Seas
5138	s5138	TV Show	Red vs. Blue	NaN	Burnie Burns, Jason Saldaña, Gustavo Sorola, G...	United States	NaT	NR	Seas
6066	s6066	TV Show	The Adventures of Figaro Pho	NaN	Luke Jurevicius, Craig Behenna, Charlotte	Australia	NaT	TV-Y7	Seas



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

7787 rows x 12 columns

In [20]:

```
#Q. 13) Find all the instances where: Category is 'Movie' and Type is 'Dramas'  
df[(df["Category"] == "Movie") & (df["Type"] == "Dramas") | (df["Category"] ==
```

Out[20]:

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	2020-01-01	PG-13	123 min
7	s8	Movie	187	Kevin Reynolds	Samuel L. Jackson, John Heard, Kelly Rowan, Cl...	United States	2019-11-01	R	119 min
111	s112	TV Show	44 Cats	NaN	Sarah Natochenny, Suzy Myers, Simona Berman, E...	Italy	2020-10-01	TV-Y7	2 Seasons
170	s171	Movie	A Family Man	Mark Williams	Gerard Butler, Gretchen Mol, Alison Brie, Will...	Canada, United States	2019-12-15	R	110 min
232	s233	Movie	A Stoning in Fulham County	Larry Elikann	Ken Olin, Jill Eikenberry, Maureen Mueller, Gr...	United States	2011-10-01	TV-14	95 min
...	...	...	...	...	...	...	...	...	
7668	s7667	TV Show	World of Winx	Iginio Straffi	Haven Paschall, Alysha Deslorieux, Jessica Paq...	Italy, United States	2017-06-16	TV-Y7	2 Seasons
7717	s7716	TV Show	Yoko	NaN	Eileen Stevens, Alyson Leigh Rosenfeld, Sarah ...	NaN	2018-06-23	TV-Y	1 Season
7719	s7718	TV Show	YOM	NaN	Sairaj, Devyani Dagaonkar, Ketan	NaN	2018-06-07	TV-Y7	1 Season

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration
					Singh, Mayur ...				
7758	s7757	TV Show	Z4	NaN	Apollo Costa, Gabriel Santana, Matheus Lustosa...	Brazil	2019-02-22	TV-PG	2 Seasons
7761	s7760	TV Show	Zak Storm	NaN	Michael Johnston, Jessica Gee- George, Christin...	United States, France, South Korea, Indonesia	2018-09-13	TV-Y7	3 Seasons