```
# importing the libraries
import pandas as pd
import matplotlib.pyplot as mlt
import seaborn as sns
import numpy as np
%matplotlib inline
```

```
# importing the dataset
df=pd.read_csv('/content/8. Netflix Dataset sk.csv')
```

```
#to show top 5 records of the data
df.head()
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Descript
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 fu where the inhab islar
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	Af devasta earthquake Mexico (
2	s3	Movie	23:59	Gilbert	Tedd Chan, Stella Chung,	Singapore	December 20,	R	78 min	Horror Movies,	When an a recruit is fc

```
#to show bottom 5 records of the data
df.tail()
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Des
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	Zoz
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	bu
7786	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	TV-MA	44 min	Documentaries, International Movies, Music & M...	doc Sou

```
# to show rows and columns of the data
df.shape
```

```
    (7789, 11)
```

```
#to show the number of total values(elements) in the dataset
df.size
```

    85679

```
#to show each column name
df.columns
```

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

```
#to show data type of each columns
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
```

```
#to show indexes,columns,data-types of each columns,memory at once
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
```

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

```
#to show row wise and detect the duplicated records.
df[df.duplicated()]
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Descript
6300	s684	Movie	Backfire	Dave Patten	Black Deniro, Byron "Squally" Vinson, Dominic ... Seun Ajayi	United States	April 5, 2019	TV-MA	97 min	Dramas, Independent Movies, Thrillers	When woul rob accider kil

```
# to remove duplicate row permanently
df.drop_duplicates(inplace=True)
```

```
# again code rum to see if duplicates are removed permanently or not
df[df.duplicated()]
```
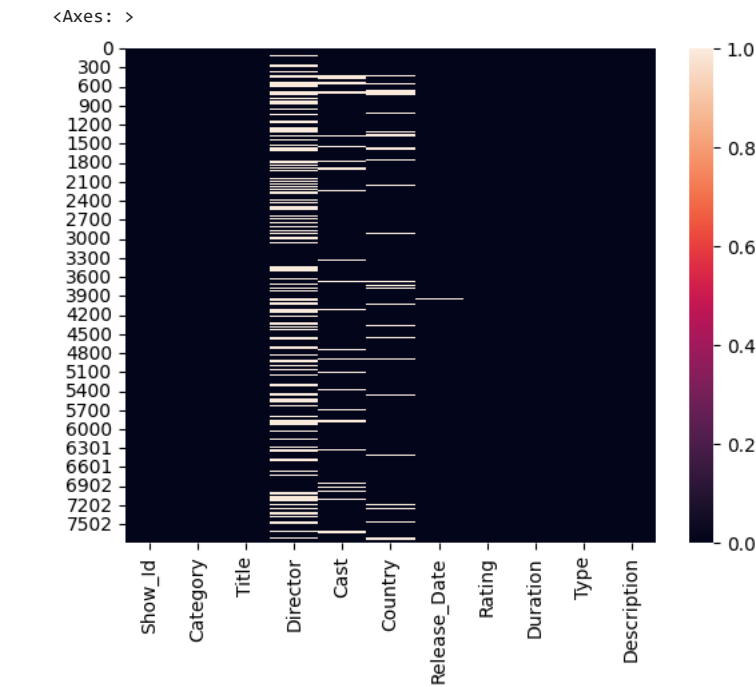
Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description

Is there any null value present in that column? Show with the heat map.

```
# to show the null values present
df.isnull().sum()
```

```
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
```

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

<Axes: >

fOR 'HOUSE OF CARDS' WHAT IS THE SHOW ID AND WHO IS THE DIRECTOR OF THE SHOW?

```
#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	United States	November 2, 2018	TV-MA	6 Seasons	TV Dramas, TV Thrillers	A ruthless politician will stop at nothing to ...

ANOTHER METHOD USING STR CONTAINS

```
# to show all records of a particular string in any column
df[df['Title'].str.contains('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	United States	November 2, 2018	TV-MA	6 Seasons	TV Dramas, TV Thrillers	A ruthless politician will stop at nothing to ...

IN WHICH YEAR HIGHEST NUMBER OF TV SHOWS AND MOVIES WERE RELEASED? SHOW WITH BAR GRAPH

```
# create new column Date_N
df['Date_N']=pd.to_datetime(df['Release_Date'])
```

```
# to count the occurence of all individual years in date columns
df['Date_N'].dt.year.value_counts()
```

```
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
```

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

```
<Axes: >
```
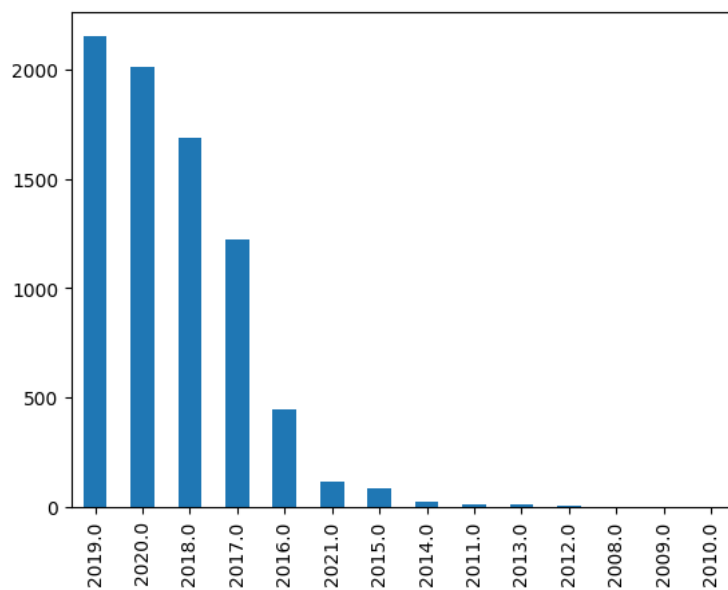


```python
df.groupby('Category').Category.count()
```

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

```python
df['Year']=df['Date_N'].dt.year
```

```python
#show all movies that were released in year 2020
df[(df['Category']=='Movie') & (df['Year']==2020)]
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type
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
					Omar Epps, Kate Cosgrove					Independent Thrillers

Show only the titles of all the tv shows that were released in India only

```
df[(df['Category']=='TV Show') & (df['Country']=='India')]
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	De
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	g
132	s133	TV Show	7 (Seven)	Nizar Shafi	Rahman, Havish, Regina Cassandra, Nandita Swet...	India	July 30, 2019	TV-14	1 Season	TV Shows	h
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...	er s
364	s365	TV Show	Akbar Birbal	NaN	Kiku Sharda, Vishal Kotian, Delnaaz Irani	India	March 31, 2020	TV-G	1 Season	Kids' TV, TV Comedies, TV Dramas	F
533	s534	TV Show	Anjaan: Rural Myths	NaN	NaN	India	December 15, 2018	TV-14	1 Season	International TV Shows, TV Horror, TV Thrillers	Sp
...	...	...	...	...	...	...	...	...	...	...	
6249	s6249	TV Show	The Creative Indians	NaN	NaN	India	August 24, 2020	TV-MA	2 Seasons	Docuseries, International TV Shows	co
6400	s6399	TV Show	The Golden Years with Javed Akhtar	NaN	NaN	India	June 1, 2017	TV-G	1 Season	Docuseries, International TV Shows	Ja

```
# To see only title use ['Title at the end]
df[(df['Category']=='TV Show') & (df['Country']=='India')]['Title']
```

```
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
```

SHOW TOP 10 DIRECTORS , WHO GAVE THE HIGHEST NUMBER OF TV SHOWS AND MOVIES TO NETFLIX?

```
# head(10) is used to find out top 10 directors
df['Director'].value_counts().head(10)
```

```
    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
```

SHOW ALL THE RECORDS WHERE CATEGORY IS MOVIE AND TYPE IS COMEDIAN OR COUNTRY IS UNITED KINGDOM

```
df[(df['Category']=='Movies') & (df['Type']=='Comedian') | (df['Country']=='United Kingdom')]
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type
19	s20	Movie	'89	NaN	Lee Dixon, Ian Wright, Paul Merson	United Kingdom	May 16, 2018	TV-PG	87 min	Sports Movies
58	s59	TV Show	100% Hotter	NaN	Daniel Palmer, Melissa Sophia, Karen Williams,...	United Kingdom	November 1, 2019	TV-14	1 Season	British TV Shows, International TV Shows, Real...
82	s83	Movie	2036 Origin Unknown	Hasraf Dulull	Katee Sackhoff, Ray Fearon, Julie Cox, Steven ...	United Kingdom	December 20, 2018	TV-14	95 min	Sci-Fi & Fantasy
85	s86	TV Show	21 Again	NaN	Laura Morgan	United Kingdom	July 1, 2020	TV-MA	1 Season	British TV Shows, Reality TV
93	s94	Movie	27: Gone Too Soon	Simon Napier-Bell	Janis Joplin, Jimi Hendrix, Amy Winehouse, Jim...	United Kingdom	May 1, 2018	TV-MA	70 min	Documentaries
...	...	...	...	...	...	...	...	...	...	...
7670	s7669	TV Show	World War II in Colour	NaN	Robert Powell	United Kingdom	August 1, 2017	TV-MA	1 Season	British TV Shows, Docuseries, International TV...
7671	s7670	TV Show	World's Busiest Cities	NaN	Anita Rani, Ade Adepitan, Dan Snow	United Kingdom	February 1, 2019	TV-PG	1 Season	British TV Shows, Docuseries
7688	s7687	Movie	XV: Beyond the Tri...	Pierre Deschamps	NaN	United Kingdom	March 18, 2020	TV-14	91 min	Documentaries, Sports Movies

IN HOW MANY MOVIES/SHOWS TOM CRUISE WAS CAST?

```
df.head(2)
```

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

```python
df[df['Cast']=='Tom Cruise']
```

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Yea
				R...								

```python
df_new=df.dropna()
```

```python
df_new[df_new['Cast'].str.contains('Tom Cruise')]
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Descri
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	Th ch h actio history
5071	s5071	Movie	Rain	Barry	Dustin Hoffman, Tom Cruise	United	July 1, 2019	R	134 min	Classic Movies	A fast-t yu for

WHAT ARE THE DIFFERENT RATINGS DEFINED BY NETFLIX?

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

```
14
```

```python
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)
```

1-HOW MANY MOVIES GOT THE 'TV-14' IN CANADA?

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

```
(1272, 13)
```

```python
df[(df['Category']=='Movies') & (df['Rating']=='TV-14') & (df['Country']=='Canada')].shape
```

```
(0, 13)
```

HOW MANY TV SHOWS GOT THE 'R' RATING AFTER YEAR 2018?

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

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

WHAT IS THE MAXIMUM DURATION OF A MOVIE/SHOW ON NETFLIX?

```python
df.Duration.unique()
```

```
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)
```

```
df.Duration.dtypes
```

```
    dtype('O')
```

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

```
df['Minute'].max()
```

```
    '99'
```

```
df['Minute'].min()
```

```
    '1'
```

```
df['Minute'].mean()
```

```
    inf
```

```
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]
    Year                float64
    Minute               object
    Unit                 object
    dtype: object
```

WHICH INDIVIDUAL COUNTRY HAS HIGHEST NUMBER OF TV SHOWS?

```
df_tvshow=df[df['Category']=='TV Show']
```

```
df_tvshow.Country.value_counts()
```

```
    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
```

```
df_tvshow.Country.value_counts().head(1)
```

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

HOW CAN WE SORT THE DATASET BY YEAR?

```
df.sort_values(by='Year').head(2)
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Descri
7115	s7114	Movie	To and From New York	Sorin Dan Mihalcescu	Barbara King, Shaana Diya, John Krisiukenas, Y...	United States	January 1, 2008	TV-MA	81 min	Dramas, Independent Movies, Thrillers	cov story York
1765	s1766	TV Show	Dinner for Five	NaN	NaN	United States	February 4, 2008	TV-MA	1 Season	Stand-Up Comedy & Talk Shows	episo cel join ho

```
df.sort_values(by='Year',ascending=False).head(2)
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Descri
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 ec detecti
5919	s5919	Movie	Surf's Up	Ash Brannon, Chris Buck	Shia LaBeouf, Jeff Bridges, Zooey Deschanel	United States	January 1, 2021	PG	86 min	Children & Family Movies, Comedies, Sports	This nom ani comedy

FIND ALL THE INSTANCE WHERE: CATEGORY IS MOVIE AND TYPE IS DRAMA OR CATEGORY IS TV SHOWS AND TYPE IS KIDS TV

```
df[(df['Category']=='Movie') & (df['Type']=='Dramas') | (df['Category']=='TV shows') & (df['Type']=='Kids TV')] .head(1)
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Dat
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar	United States	January 1, 2020	PG-13	123 min	Dramas	A brilliant group of students become card-coun...	2(0'

✓   0s   completed at 09:39   ● ✕