Netflix Data Analysis Using Python

The Netflx data set has the information about the Tv Shows & Movies available on Netflix till 2021.

The Data set available from Flexible which is a Third Party Netflix which engine , and available on Kaggle dataset for free.

Import Library

In [1]:
```
import pandas as pd
```

In [2]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import seaborn as sns
```

```
C:\Users\Syed Arif\anaconda3\lib\site-packages\scipy\__init__.py:146: UserWar
ning: A NumPy version >=1.16.5 and <1.23.0 is required for this version of Sc
iPy (detected version 1.25.1
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

Uploading Csv fle

In [3]:
```
df = pd.read_csv(r"C:\Users\Syed Arif\Downloads\8. Netflix Dataset.csv")
```

Data Preprocessing

.head()

head is used show to the By default = 5 rows in the dataset

In [4]:
```
df.head()
```

Out[4]:

	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	In T
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	In
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 min	In
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	80 min	In M
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	

.tail()

tail is used to show rows by Descending order

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

Out[5]:

	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 Seasor
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

.shape

It show the total no of rows & Column in the dataset

In [6]:
```
df.shape
```

Out[6]:
```
(7789, 11)
```

.Columns

It show the no of each Column

In [7]:
```
df.columns
```

Out[7]:
```
Index(['Show_Id', 'Category', 'Title', 'Director', 'Cast', 'Country',
       'Release_Date', 'Rating', 'Duration', 'Type', 'Description'],
      dtype='object')
```

.dtypes

This Attribute show the data type of each column

In [8]:
```
df.dtypes
```

Out[8]:
```
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
```

.unique()

In a column, It show the unique value of specific column.

In [9]:
```
df["Country"].unique()
```

```
       'Israel, United States', 'United States, Mexico',
       'Uruguay, Argentina, Spain', 'Singapore, France',
       'United Kingdom, United States, France, Germany',
       'Turkey, United States', 'Bulgaria, United States',
       'Australia, France', 'Hong Kong, Iceland, United States',
       'United Arab Emirates', 'United States, Chile',
       'Germany, France, Russia', 'Mauritius, South Africa',
       'United States, Japan', 'Lebanon', 'United States, Bulgaria',
       'Colombia', 'Uruguay, Argentina', 'Egypt, Algeria',
       'France, Egypt', 'Uruguay', 'Soviet Union, India',
       'Sweden, United States', 'South Africa', 'Malaysia',
       'Ireland, United Kingdom, United States', 'Spain, Italy',
       'United Kingdom, France, Germany',
       'United States, Germany, Canada', 'United States, India',
       'Japan, United States', 'Denmark, United States',
       'South Africa, United States', 'Canada, Luxembourg',
       'Serbia, United States', 'Canada, Nigeria',
       'Iceland, Sweden, Belgium', 'Ireland, Canada',
       'United States, Italy', 'Finland', 'India, Germany',
```

.nuique()

It will show the total no of unque value from whole data frame

In [10]:
```
df.nunique()
```

Out[10]:
```
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
```

.describe()

It show the Count, mean , median etc

In [11]:
```
df.describe()
```

Out[11]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Du
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	S
freq	2	5379	2	18	18	2556	118	2865	

.value_counts

It Shows all the unique values with their count

In [12]:
```
df["Country"].value_counts()
```

Out[12]:
```
United States                                                 2556
India                                                          923
United Kingdom                                                 397
Japan                                                          226
South Korea                                                    183
                                                              ...
Russia, United States, China                                     1
Italy, Switzerland, France, Germany                              1
United States, United Kingdom, Canada                            1
United States, United Kingdom, Japan                             1
Sweden, Czech Republic, United Kingdom, Denmark, Netherlands     1
Name: Country, Length: 681, dtype: int64
```

.isnull()

It shows the how many null values

In [13]:
```
df.isnull()
```

Out[13]:

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

7789 rows × 11 columns

.info()

To Show Data type of each column

In [14]:
```
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
```

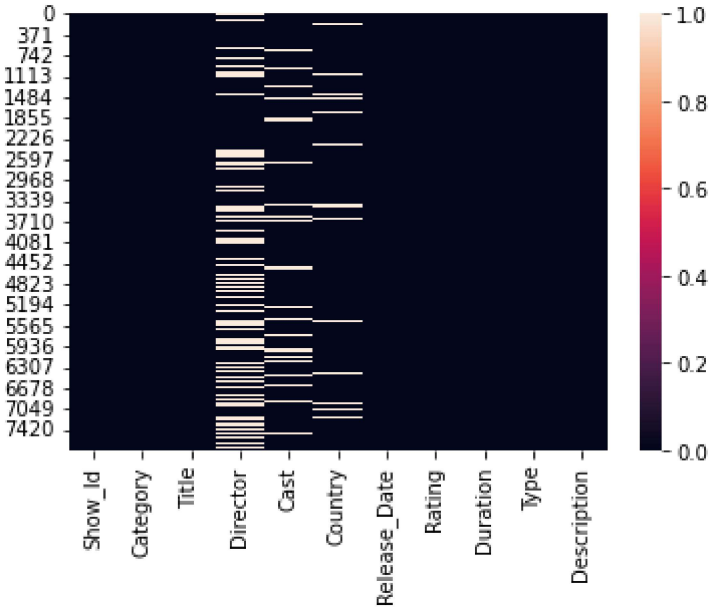
Is there any Null value present in any Column ? Show with heatmap

In [15]:
```
df.isnull().sum()
```

Out[15]:
```
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 [16]:
```
sns.heatmap(df.isnull())
```

Out[16]: <AxesSubplot:>



For "Zozo" ; What is the Show Case Id and Who is the the Director of this Show ?

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

Out[17]:

	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	In T
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	In
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 min	In
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	80 min	In M
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	

In [56]:
```
df[df["Title"].str.contains("Zozo")]
```

Out[56]:

	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 min	I

In which year Heighest Number of Tv Shows and Movies were released ?

In [57]:
```
df.head()
```

Out[57]:

	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	In T
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	In
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 min	In
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	80 min	In M
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	

In [58]:
```
df.dtypes
```

Out[58]:
```
Show_Id                 object
Category                object
Title                   object
Director                object
Cast                    object
Country                 object
Release_Date            object
Rating                  object
Duration                object
Type                    object
Description             object
N_Date          datetime64[ns]
Year                   float64
dtype: object
```

First Convert the data type of Column Release_Date ? Using to_datetime

In [59]:
```
df["N_Date"] = pd.to_datetime(df['Release_Date'])
```

In [60]:
```
df
```

Out[60]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Durati
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	Seasc
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 r
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 r
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	80 r
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 r
...	...	...	...	...	...	...	...	...	
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 r
7785	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	TV-14	111 r
7786	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	TV-MA	44 r
7787	s7786	TV Show	Zumbo's Just Desserts	NaN	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	TV-PG	Seas

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Durati
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 r

7789 rows × 13 columns

Pick Only Years

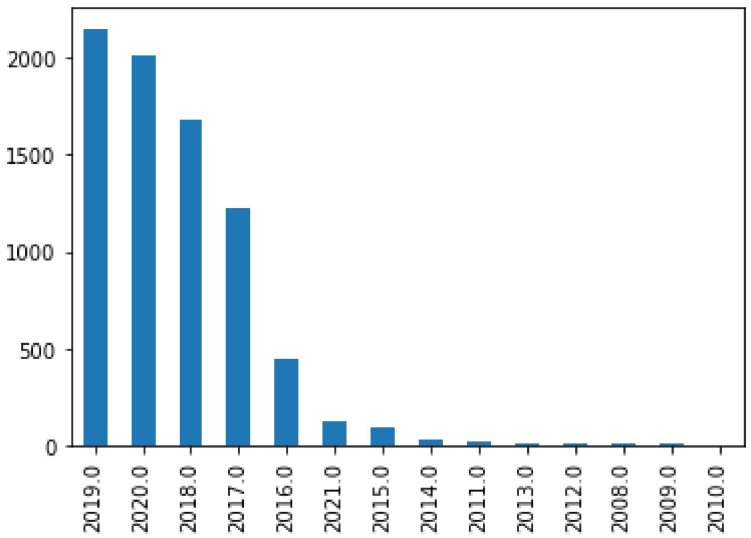
In [61]:
```
df["N_Date"].dt.year.value_counts()
```

Out[61]:
```
2019.0    2154
2020.0    2010
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: N_Date, dtype: int64
```

Bar Graph

In [62]:
```
df["N_Date"].dt.year.value_counts().plot(kind = "bar")
```

Out[62]:
```
<AxesSubplot:>
```



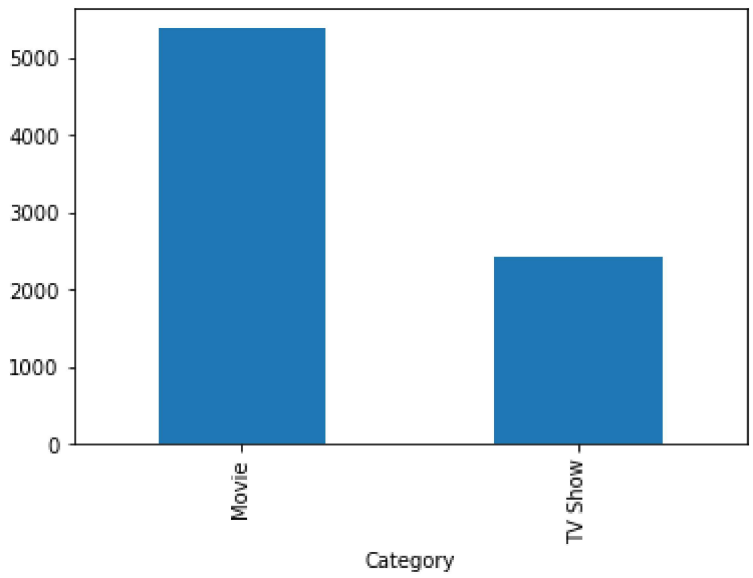
How many Movies & Tv shows are inthe dataset ? Show With bar grapp

In [63]:
```
df.groupby("Category").Category.count()
```

Out[63]:
```
Category
Movie      5379
TV Show    2410
Name: Category, dtype: int64
```

In [64]:
```
df.groupby("Category").Category.count().plot(kind = "bar")
```

Out[64]:
```
<AxesSubplot:xlabel='Category'>
```



Show all the Movies that were Released in 2010 ?

In [65]:
```
df.head()
```

Out[65]:

	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	In T
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	In
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 min	In
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	80 min	In M
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	

In [66]:
```
df["Year"] = df["N_Date"].dt.year
```

In [67]:
```
df[(df["Category"] == 'Movie' ) & (df['Year']  == 2010)]
```

Out[67]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration
3840	s3841	Movie	Mad Ron's Prevues from Hell	Jim Monaco	Nick Pawlow, Jordu Schell, Jay Kushwara, Micha...	United States	November 1, 2010	NR	84 min

Show only Titles off all Tv shows that were released in Pakistan Only ?

In [84]:
```
df[(df["Category"] == "Tv Show") & (df["Country"] == "United States")]
```

Out[84]:

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Desc

Show Top 10 Directors , Who gave the Heighest Number of Tv Shows & Movies to Netflix ?

In [89]:
```
df["Director"].value_counts().head(10)
```

Out[89]:
```
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
Robert Rodriguez           8
Name: Director, dtype: int64
```

Show all the Records, were "Category" is "Movie" and "type" is "Horror Movies" or "Country" is "United Kingdom".

In [96]:
```
df[(df["Category"] == 'Movie' ) & (df['Type']  == "Horror Movies")]
```

Out[96]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating
261	s262	Movie	A.M.I.	Rusty Nixon	Debs Howard, Philip Granger, Sam Robert Muik, ...	Canada	October 1, 2020	TV-MA
417	s418	Movie	All the Boys Love Mandy Lane	Jonathan Levine	Anson Mount, Edwin Hodge, Michael Welch, Brook...	United States	July 3, 2018	R
1551	s1552	Movie	Cult of Chucky	Don Mancini	Fiona Dourif, Michael Therriault,	United States	October 3, 2017	R

What are the different rating defined By the Netflix ?

In [97]:
```
df.head(2)
```

Out[97]:

	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	Inte TV TV
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	Inte

In [99]:
```
df.Rating.unique()
```

Out[99]:
```
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)
```

What is the maximum duration of Netflix On Tv Show ?

In [100]:
```
df.head(1)
```

Out[100]:

	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	Inter TV TV D T

In [101]:
```
df['Duration'].unique()
```

Out[101]:
```
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)
```

The Duration Column has 12 tpe values one is Int and oother is Object, Seprate Both the Values Using "str.split" Dunction

In [102]:
```
df[['Minutes' , 'Unit']] = df["Duration"].str.split( " " , expand = True)
```

In [103]:
```
df.head(1)
```

Out[103]:

	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	Inter TV TV D T

In [104]:
```
df.Minutes.max()
```

Out[104]:
```
'99'
```

How we can sort the dataset By year

In [105]:
```
df.sort_values(by = 'Year').head()
```

Out[105]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Dura
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
1765	s1766	TV Show	Dinner for Five	NaN	NaN	United States	February 4, 2008	TV-MA	Sea
5766	s5766	Movie	Splatter	Joe Dante	Corey Feldman, Tony Todd, Tara Leigh, Erin Way...	United States	November 18, 2009	TV-MA	29
3248	s3249	Movie	Just Another Love Story	Ole Bornedal	Anders W. Berthelsen, Rebecka Hemse, Nikolaj L...	Denmark	May 5, 2009	TV-MA	104
3840	s3841	Movie	Mad Ron's Prevues from Hell	Jim Monaco	Nick Pawlow, Jordu Schell, Jay Kushwara, Micha...	United States	November 1, 2010	NR	84

In [ ]: