

Netflix Dataset

Big-Data Analysis with Python



This Netflix Dataset has information about the TV Shows and Movies available on Netflix till 2021.

This Dataset is Collected Flixible which is a third-party Netflix Search Engine and available on kaggle website for free.

```
In [1]: 1 import pandas as pd  
        2 data = pd.read_csv("Netflix.csv")
```

In [56]:

1 data.head(7)

Out[56]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
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 ...	2020-08-14	2020.0
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...	2016-12-23	2016.0
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 army recruit is found dead, his fellow...	2018-12-20	2018.0
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...	In a postapocalyptic world, rag-doll robots hi...	2017-11-16	2017.0
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	Dramas	Abrilliant group of students become card-coun...	2020-01-01	2020.0
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	Agenetics professor experiments with a treatm...	2017-07-01	2017.0
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	After an awful accident, a couple admitted to ...	2020-06-01	2020.0

Getting some information about the dataset:

1.Head()

In [57]:

1data.head(2)

Out[57]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
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 ...	2020-08-14	2020.0
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...	2016-12-23	2016.0

2.Tail()

In [58]:

1data.tail(2)

Out[58]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
7787	s7786	TV Show	Zumbo's Just Desserts	NaN	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	TV-PG	1 Season	International TV Shows, Reality TV	Dessert wizard Adriano Zumbo looks for the nex...	2020-10-31	2020.0
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 min	Documentaries, Music & Musicals	This documentary delves into the mystique behi...	2020-03-01	2020.0

3.Shape

In [5]:

1data.shape

Out[5]: (7789, 11)

4.Size

```
In [6]: 1 data.size
```

```
Out[6]: 85679
```

5.Columns

```
In [7]: 1 data.columns
```

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

6.dtypes

```
In [8]: 1 data.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
```

7.Info()

```
In [9]: 1 data.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 Record in this dataset ? If yes, then remove the duplicate records.

Duplicate()

```
In [10]: 1 data[data.duplicated()]
```

Out[10]:

	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 ...	United States	April 5, 2019	TV-MA	97 min	Dramas, Independent Movies, Thrillers	When two would-be robbers accidentally kill a ...
6622	s6621	Movie	The Lost Okoroshi	Abba T. Makama	Seun Ajayi, Judith Audu, Tope Tedela, Ifu Enna...	Nigeria	September 4, 2020	TV-MA	94 min	Comedies, Dramas, Independent Movies	Adisillusioned security guard transforms into...

```
In [11]: 1 data.drop_duplicates(inplace = True)
```

```
In [12]: 1 data[data.duplicated()]
```

```
Out[12]:
```

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

```
In [13]: 1 data.shape
```

```
Out[13]: (7787, 11)
```

Task. 2) Is there any Null Value present in any column ? Show with Heat-map.

IsNull()

```
In [59]: 1 data.isnull().head(2)
```

```
Out[59]:
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
0	False	False	False	True	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False

```
In [15]: 1 data.isnull().sum()
```

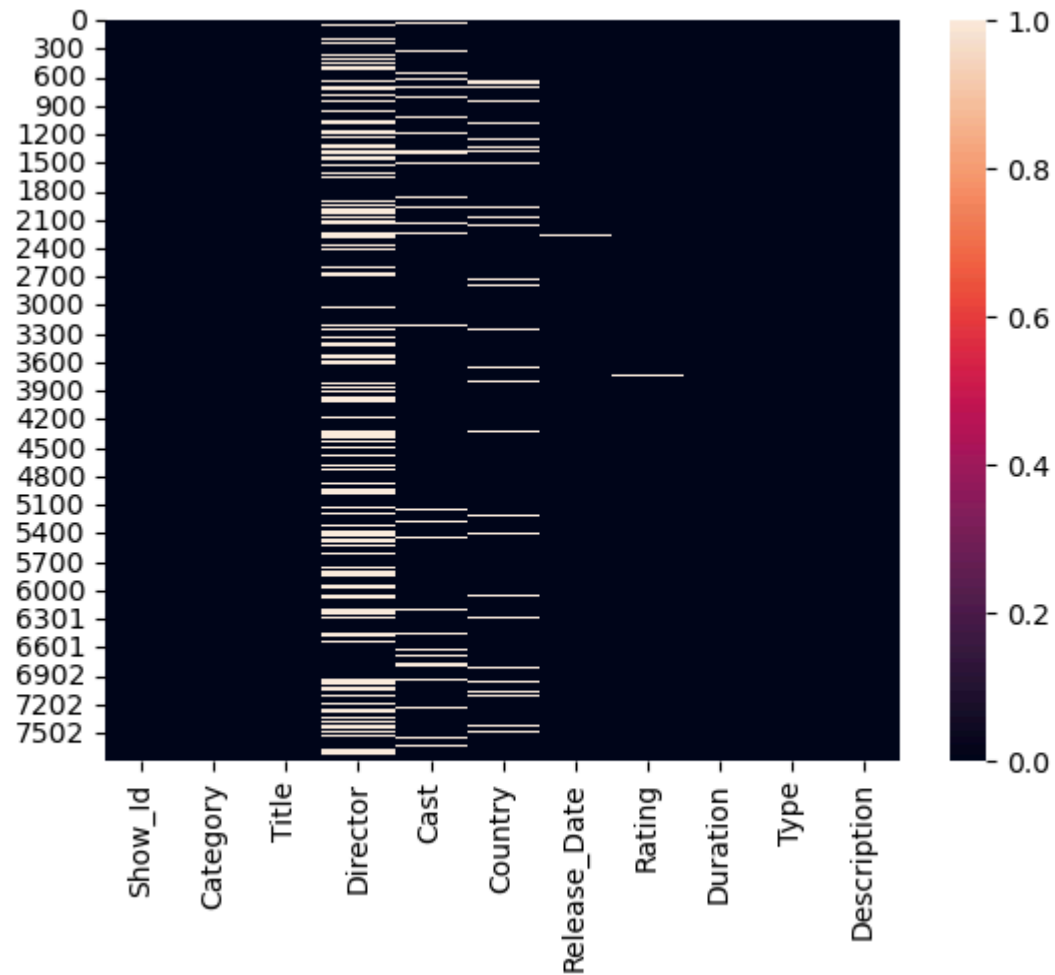
```
Out[15]: Show_Id      0
Category      0
Title         0
Director     2388
Cast         718
Country      507
Release_Date  10
Rating        7
Duration      0
Type          0
Description   0
dtype: int64
```

Seaborn Library (Heat-Map)

```
In [16]: 1 import seaborn as sns
```

```
In [17]: 1 sns.heatmap(data.isnull())
```

```
Out[17]: <Axes: >
```



Q.1. House of Cards what is the show id and who is the director the show ?

Isin()

```
In [18]: 1 data[data['Title'].isin(['House of Cards'])]
```

```
Out[18]:
```

	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, J...	Kevin Spacey, Robin Wright, Kate Mara, Corey S...	United States	November 2, 2018	TV-MA	6 Seasons	TV Dramas, TV Thrillers	Aruthless politician will stop at nothing to ...

Q.2. Which year highest numbers of TV Shows and Movies were released ? Show with Bar Graph.

To_DateTime

```
In [19]: 1 data['Date_N'] = pd.to_datetime(data['Release_Date'])
```

```
In [60]: 1 data.head(2)
```

```
Out[60]:
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
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 ...	2020-08-14	2020.0
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...	2016-12-23	2016.0

dtypes

```
In [21]: 1 data.dtypes
```

```
Out[21]: 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
```

dt.year.value_counts()

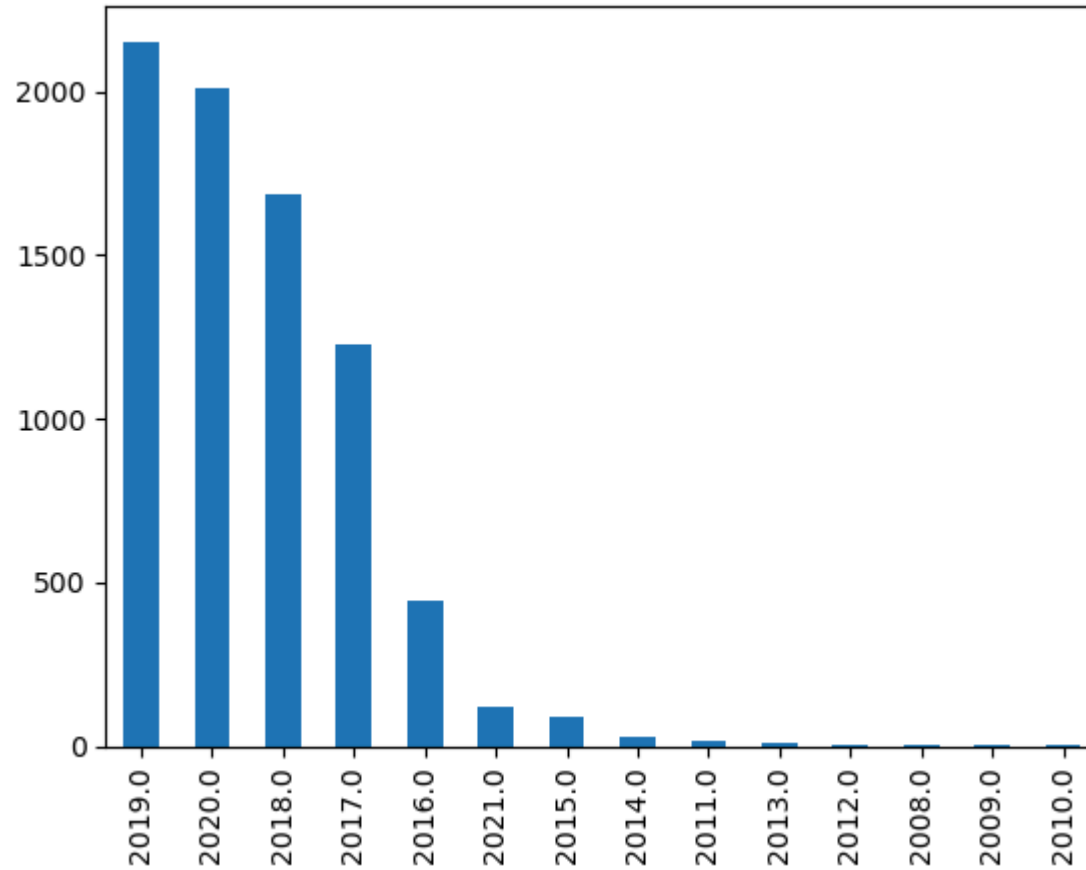
```
In [22]: 1 data['Date_N'].dt.year.value_counts()
```

```
Out[22]: 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

```
In [23]: 1 data['Date_N'].dt.year.value_counts().plot(kind='bar')
```

```
Out[23]: <Axes: >
```



Q.3. How many Movies & TV Shows Are in the Dataset ?

Groupby()

```
In [24]: 1 data.groupby('Category').Category.count()
```

```
Out[24]: Category
Movie      5377
TV Show    2410
Name: Category, dtype: int64
```

Q.4. Show all Movies that were Released in year 2000.

Crating a new Column

```
In [25]: 1 data[ 'Year' ] = data[ 'Date_N' ].dt.year
```

```
In [26]: 1 data.head(2)
```

```
Out[26]:
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
0	s1	TVShow	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 ...	2020- 08-14	2020.0
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...	2016- 12-23	2016.0

Filtering

```
In [27]: 1 data[ (data['Category'] == 'Movie') & (data['Year'] == 2000) ]
```

Out[27]:

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

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

```
In [28]: 1 data.head(2)
```

Out[28]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
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 ...	2020-08-14	2020.0
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...	2016-12-23	2016.0

```
In [29]: 1 data[(data['Category']=='TV Show') & (data['Country']=='India')] ['Title']
```

```
Out[29]: 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 [30]: 1 data['Director'].value_counts().head(10)
```

```
Out[30]: 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
```

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

```
In [62]: 1 data[(data['Category']=='Movie') & (data['Type']=='Comedies') | (data['Country']=='United Kingdom')].head(2)
```

```
Out[62]:
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
19	s20	Movie	'89	NaN	Lee Dixon, Ian Wright, Paul Merson	United Kingdom	May 16, 2018	TV-PG	87 min	Sports Movies	Mixing old footage with interviews, this is th...	2018-05-16	2018.0
33	s34	Movie	#realityhigh	Fernando Lebrija	Nesta Cooper, Kate Walsh, John Michael Higgins...	United States	September 8, 2017	TV-14	99 min	Comedies	When nerdy high schooler Dani finally attracts...	2017-09-08	2017.0

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

```
In [33]: 1 data[(data['Cast']=='Tom Cruise')]
```

```
Out[33]:
```

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

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

```
In [34]: 1 data['Rating'].nunique()
```

```
Out[34]: 14
```

```
In [35]: 1 data['Rating'].unique()
```

```
Out[35]: 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 ?

```
In [37]: 1 data[(data['Category']=='Movie') & (data['Rating']=='TV-14') & (data['Country']=='Canada')].shape
```

```
Out[37]: (11, 13)
```

Q. 9.2) How many TV Shows got the 'R' rating, after year 2018 ?

```
In [39]: 1 data[(data['Category']=='TV Show') & (data['Rating']=='R') & (data['Year']>2018)]
```

```
Out[39]:
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
6437	s6436	TV Show	The Hateful Eight: Extended Version	Quentin Tarantino	Samuel L. Jackson, Kurt Russell, Jennifer Jaso...	NaN	April 25, 2019	R	1 Season	TV Shows	Trapped at a stagecoach stop as a storm rages ...	2019- 04-25	2019.0

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

```
In [41]: 1 data['Duration'].max()
```

```
Out[41]: '99 min'
```

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

```
In [43]: 1 data_tvshow = data[data['Category']=='TV Show']
```

```
In [44]: 1 data_tvshow.head(2)
```

```
Out[44]:
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
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 ...	2020-08-14	2020.0
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	Agenetics professor experiments with a treatm...	2017-07-01	2017.0

```
In [45]: 1 data_tvshow.Country.value_counts().head(1)
```

```
Out[45]: United States    705
Name: Country, dtype: int64
```

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

```
In [46]: 1 data.sort_values(by = 'Year').head(2)
```

```
Out[46]:
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
7115	s7114	Movie	To and From New York	Sorin Dan Mihalcescu	Barbara King, Shaana Diya, John Krišiukenas, Y...	United States	January 1, 2008	TV-MA	81 min	Dramas, Independent Movies, Thrillers	While covering a story in New York City, a Sea...	2008-01-01	2008.0
1765	s1766	TV Show	Dinner for Five	NaN	NaN	United States	February 4, 2008	TV-MA	1 Season	Stand-Up Comedy & Talk Shows	In each episode, four celebrities join host Jo...	2008-02-04	2008.0

In [47]: 1 data.sort_values(by = 'Year', ascending = False).head(2)

Out[47]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
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 game is afoot for an eccentric detective w...	2021-01-01	2021.0
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 Oscar-nominated animated comedy goes behi...	2021-01-01	2021.0

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

In [48]: 1 data[(data['Category']=='Movie') & (data['Type']=='Dramas')].head(2)

Out[48]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	Dramas	Abrilliant group of students become card-coun...	2020-01-01	2020.0
7	s8	Movie	187	Kevin Reynolds	Samuel L. Jackson, John Heard, Kelly Rowan, Cl...	United States	November 1, 2019	R	119 min	Dramas	After one of his high school students attacks ...	2019-11-01	2019.0

In [49]: 1 data[(data['Category']=='TV Show') & (data['Type']=="Kids' TV")].head(2)

Out[49]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
111	s112	TV Show	44 Cats	NaN	Sarah Natochenny, Suzy Myers, Simona Berman, E...	Italy	October 1, 2020	TV-Y7	2 Seasons	Kids' TV	Paw-esome tales abound when singing furry frie...	2020-10-01	2020.0
276	s277	TV Show	Abby Hatcher	Kyran Kelly	Macy Drouin, Wyatt White, Paul Sun-Hyung Lee, ...	United States, Canada	July 1, 2020	TV-Y	1 Season	Kids' TV	Abig-hearted girl helps her Fuzzly friends wh...	2020-07-01	2020.0

In [50]:

1 data[(data['Category']=='Movie') & (data['Type']=='Dramas') | (data['Category']=='TV Show') & (data['Type']=="Kids' TV

Out[50]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	Date_N	Year
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	Dramas	Abrilliant group of students become card-coun...	2020-01-01	2020.0
7	s8	Movie	187	Kevin Reynolds	Samuel L. Jackson, John Heard, Kelly Rowan, Cl...	United States	November 1, 2019	R	119 min	Dramas	After one of his high school students attacks ...	2019-11-01	2019.0
111	s112	TV Show	44 Cats	NaN	Sarah Natochenny, Suzy Myers, Simona Berman, E...	Italy	October 1, 2020	TV-Y7	2 Seasons	Kids' TV	Paw-esome tales abound when singing furry frie...	2020-10-01	2020.0
170	s171	Movie	AFamily Man	Mark Williams	Gerard Butler, Gretchen Mol, Alison Brie, Will...	Canada, United States	December 15, 2019	R	110 min	Dramas	Aruthless corporate headhunter battles his ri...	2019-12-15	2019.0
232	s233	Movie	AStoning in Fulham County	Larry Elikann	Ken Olin, Jill Eikenberry, Maureen Mueller, Gr...	United States	October 1, 2011	TV-14	95 min	Dramas	After reckless teens kill an Amish child, a pr...	2011-10-01	2011.0