

netflix-data-analysis

May 5, 2023

```
[1]: import warnings
warnings.filterwarnings('ignore')
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
[2]: df = pd.read_csv(r"Netflix dataset.csv")
df.head()
```

```
[2]: Show_Id Category Title Director \
0 s1 TV Show 3% NaN
1 s2 Movie 07:19 Jorge Michel Grau
2 s3 Movie 23:59 Gilbert Chan
3 s4 Movie 9 Shane Acker
4 s5 Movie 21 Robert Luketic

Cast Country \
0 João Miguel, Bianca Comparato, Michel Gomes, R... Brazil
1 Demián Bichir, Héctor Bonilla, Oscar Serrano, ... Mexico
2 Tedd Chan, Stella Chung, Henley Hii, Lawrence ... Singapore
3 Elijah Wood, John C. Reilly, Jennifer Connelly... United States
4 Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar... United States

Release_Date Rating Duration \
0 August 14, 2020 TV-MA 4 Seasons
1 December 23, 2016 TV-MA 93 min
2 December 20, 2018 R 78 min
3 November 16, 2017 PG-13 80 min
4 January 1, 2020 PG-13 123 min

Type \
0 International TV Shows, TV Dramas, TV Sci-Fi &...
1 Dramas, International Movies
2 Horror Movies, International Movies
3 Action & Adventure, Independent Movies, Sci-Fi...
```

	Description
0	In a future where the elite inhabit an island ...
1	After a devastating earthquake hits Mexico Cit...
2	When an army recruit is found dead, his fellow...
3	In a postapocalyptic world, rag-doll robots hi...
4	A brilliant group of students become card-coun...

```
[3]: df.shape
```

```
[3]: (7789, 11)
```

```
[4]: 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
```

Q1. Is there is a duplicate records in this dataset ? if yes, then remove them

```
[5]: df.duplicated().sum()
```

```
[5]: 2
```

```
[6]: df.drop_duplicates(inplace=True)
```

```
[7]: df.shape
```

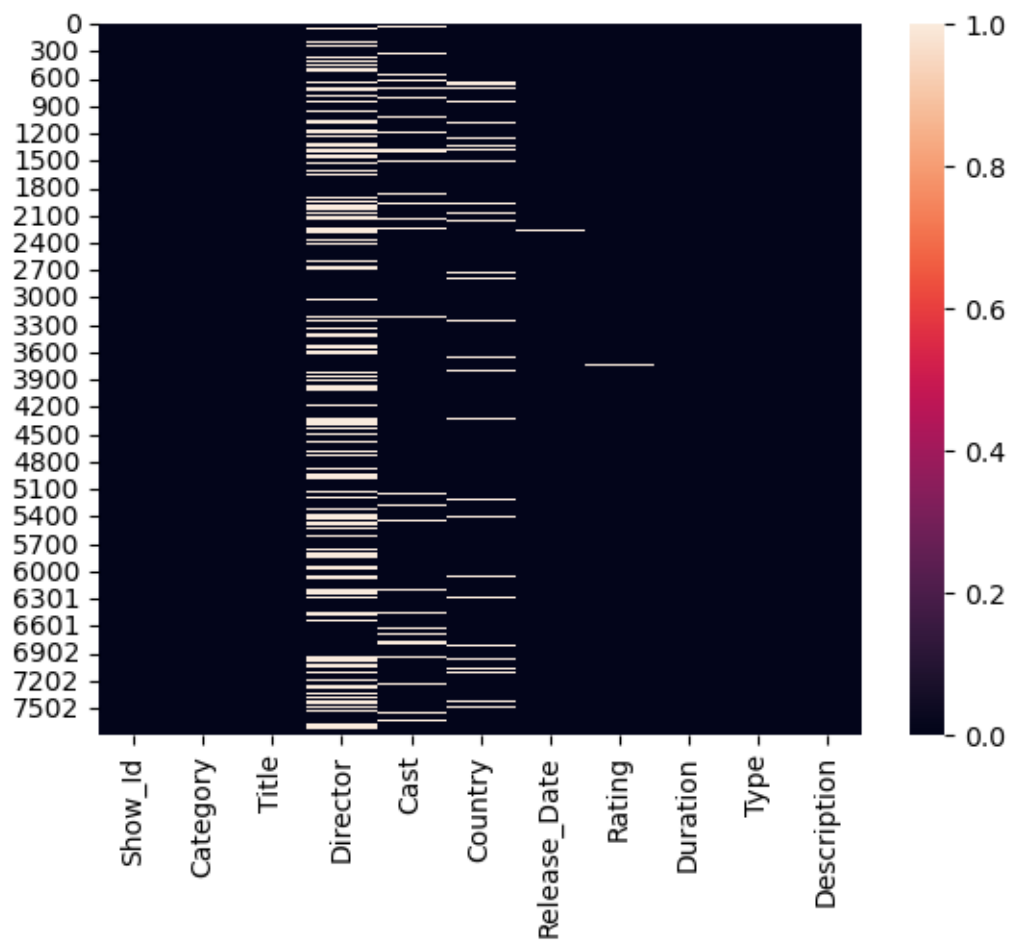
```
[7]: (7787, 11)
```

Q2. Is there is a null values in this data ? If yes then show it in the heatmap

```
[8]: df.isnull().sum()
```

```
[8]: 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
```

```
[9]: sns.heatmap(df.isnull())
     plt.show()
```



Q3. For 'House of cards' movie what is the show_id and who is the director of this movie ?

```
[10]: df.loc[df["Title"]=="House of Cards", ["Title","Show_Id","Director"]]
```

```
[10]:          Title Show_Id \
2832  House of Cards    s2833

          Director
2832  Robin Wright, David Fincher, Gerald McRaney, J...
```

Q4. In which year highest number of TV shows and movies released ? show it in the bar chart

```
[11]: df.dtypes # date column is in str format required to be change in date
```

```
[11]: 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
```

```
[12]: df.Release_Date = pd.to_datetime(df.Release_Date)
```

```
[13]: df.Release_Date.dtype
```

```
[13]: dtype('<M8[ns]')
```

```
[14]: df.groupby(df.Release_Date.dt.year)["Show_Id"].count().
      ↪sort_values(ascending=False)
      # in 2019 highest number of movies and web series are released
```

```
[14]: Release_Date
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
```

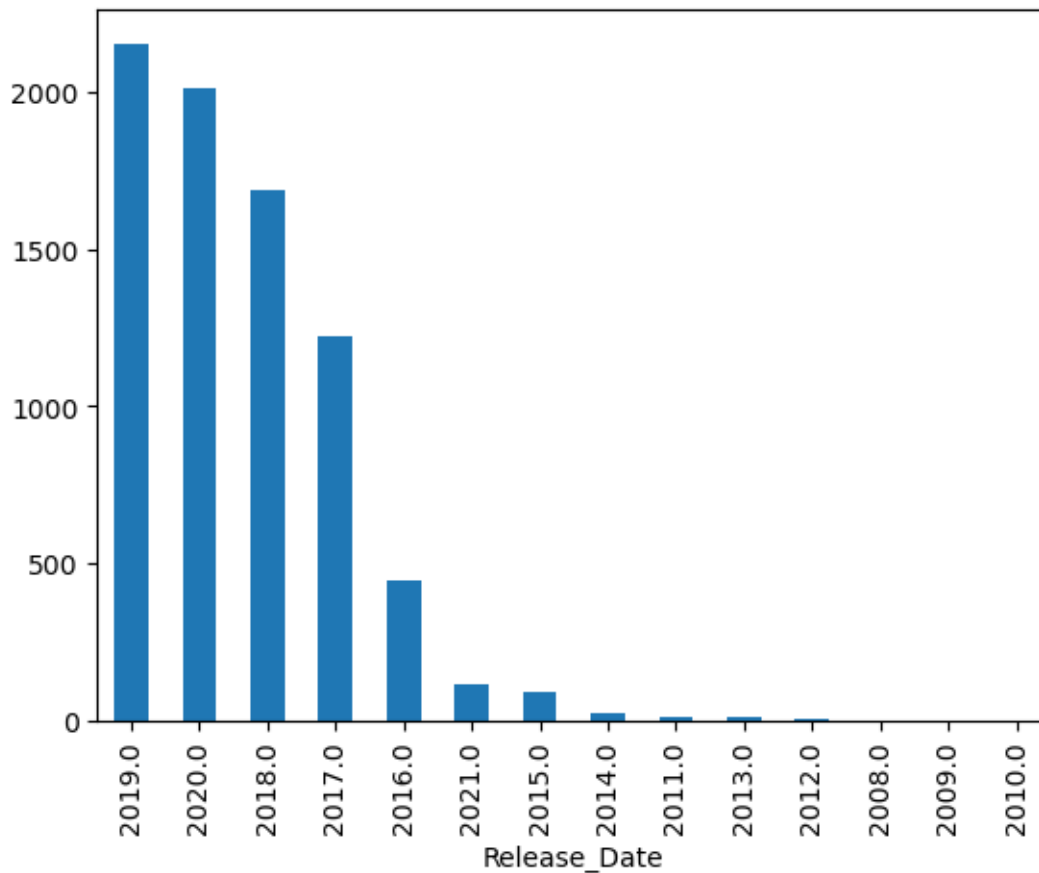
```

2013.0      11
2012.0       3
2008.0       2
2009.0       2
2010.0       1
Name: Show_Id, dtype: int64

```

```
[15]: df.groupby(df.Release_Date.dt.year)["Show_Id"].count().sort_values(
      ascending=False).plot.bar()
```

```
[15]: <AxesSubplot:xlabel='Release_Date'>
```



Q5. How many Movies and TV series are in the dataset ? Show it in bar chart

```
[16]: df["Category"].value_counts()
```

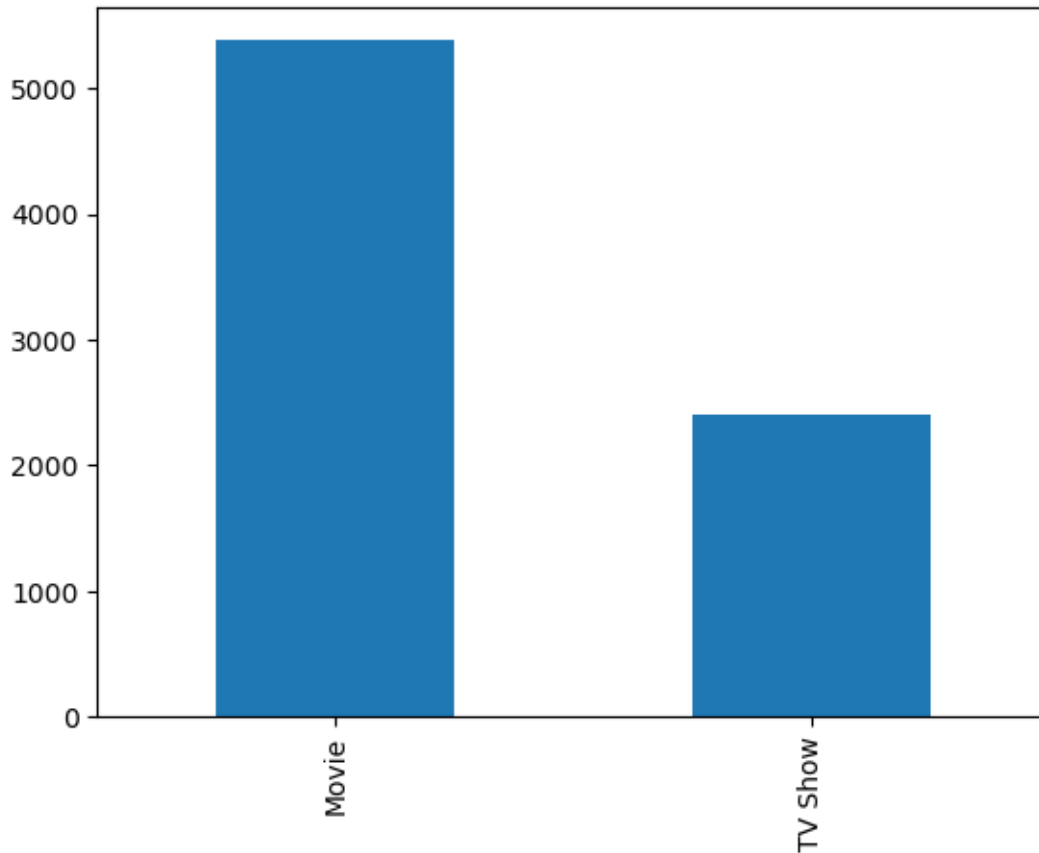
```

[16]: Movie      5377
      TV Show    2410
      Name: Category, dtype: int64

```

```
[17]: df["Category"].value_counts().plot.bar()
      # Movies are double as compare to webseries because web series trend has emerged
      # in recent 2-3 years
```

```
[17]: <AxesSubplot:>
```



Q6. How all the 'movies' are made in year 2020 ?

```
[18]: len(df[(df["Category"]=="Movie") & (df["Release_Date"].dt.year==2020)])
```

```
[18]: 1312
```

```
[19]: print("Number of movies made in year 2020 ---->>>",
      ↪len(df[(df["Category"]=="Movie") & (df["Release_Date"].dt.year==2020)]))
```

Number of movies made in year 2020 ---->>> 1312

Q7. How many Titles are released in India only ?

```
[20]: len(df[df["Country"]=="India"])
```

[20]: 923

```
[21]: print("Number of Titles released in India are_
↳----->>>", len(df[df["Country"]=="India"]))
```

Number of Titles released in India are ----->>> 923

Q8. Top 10 directors who gave highest number of movies or web series

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

```
[22]: 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
```

Q9. Show the records where 'category is movie' and 'type is comedy' or 'country is UK'

```
[23]: df[(df.Category=="Movie") & (df["Type"].str.contains("Comedies", case=False) |
      (df["Country"]=="United Kingdom"))].head(15)
```

```
[23]: Show_Id Category Title Director \
18 s19 Movie 15-Aug Swapnaneel Jayakar
19 s20 Movie '89 NaN
33 s34 Movie #realityhigh Fernando Lebrija
34 s35 Movie #Roxy Michael Kennedy
36 s37 Movie #Selfie Cristina Jacob
37 s38 Movie #Selfie 69 Cristina Jacob
39 s40 Movie ¡Ay, mi madre! Frank Ariza
40 s41 Movie Çarsı Pazar Muharrem Gülmez
42 s43 Movie Çok Filim Hareketler Bunlar Ozan Açıktan
48 s49 Movie 10 Days in Sun City Adze Ugah
49 s50 Movie 10 jours en or Nicolas Brossette
72 s73 Movie 17 Again Burr Steers
77 s78 Movie 2 Alone in Paris Ramzy Bedia, Éric Judor
78 s79 Movie 2 States Abhishek Varman
82 s83 Movie 2036 Origin Unknown Hasraf Dulull

Cast Country \
18 Rahul Pethe, Mrunmayee Deshpande, Adinath Koth...
```

19	Lee Dixon, Ian Wright, Paul Merson	United Kingdom
33	Nesta Cooper, Kate Walsh, John Michael Higgins...	United States
34	Jake Short, Sarah Fisher, Booboo Stewart, Dann...	Canada
36	Flavia Hojda, Crina Semciuc, Olimpia Melinte, ...	Romania
37	Maia Morgenstern, Olimpia Melinte, Crina Semci...	Romania
39	Estefanía de los Santos, Secun de la Rosa, Ter...	Spain
40	Erdem Yener, Ayhan Taş, Emin Olcay, Muharrem G...	Turkey
42	Ayça Erturan, Aydan Taş, Ayşegül Akdemir, Burc...	Turkey
48	Ayo Makun, Adesua Etomi, Richard Mofe-Damijo, ...	South Africa, Nigeria
49	Franck Dubosc, Claude Rich, Marie Kremer, Math...	France
72	Zac Efron, Leslie Mann, Matthew Perry, Thomas ...	United States
77	Ramzy Bedia, Éric Judor, Benoît Magimel, Krist...	France
78	Alia Bhatt, Arjun Kapoor, Ronit Roy, Amrita Si...	India
82	Katee Sackhoff, Ray Fearon, Julie Cox, Steven ...	United Kingdom

	Release_Date	Rating	Duration \
18	2019-03-29	TV-14	124 min
19	2018-05-16	TV-PG	87 min
33	2017-09-08	TV-14	99 min
34	2019-04-10	TV-14	105 min
36	2019-06-01	TV-MA	125 min
37	2019-06-01	TV-MA	119 min
39	2019-07-19	TV-MA	81 min
40	2017-03-10	TV-14	97 min
42	2017-03-10	TV-MA	99 min
48	2019-10-18	TV-14	87 min
49	2017-07-01	TV-14	97 min
72	2021-01-01	PG-13	102 min
77	2020-06-01	TV-MA	97 min
78	2018-08-04	TV-PG	143 min
82	2018-12-20	TV-14	95 min

	Type \
18	Comedies, Dramas, Independent Movies
19	Sports Movies
33	Comedies
34	Comedies, Romantic Movies
36	Comedies, Dramas, International Movies
37	Comedies, Dramas, International Movies
39	Comedies, International Movies
40	Comedies, International Movies
42	Comedies, International Movies
48	Comedies, International Movies, Romantic Movies
49	Comedies, Dramas, International Movies
72	Comedies
77	Comedies, International Movies
78	Comedies, Dramas, International Movies

	Description
18	On India's Independence Day, a zany mishap in ...
19	Mixing old footage with interviews, this is th...
33	When nerdy high schooler Dani finally attracts...
34	A teenage hacker with a huge nose helps a cool...
36	Two days before their final exams, three teen ...
37	After a painful breakup, a trio of party-lovin...
39	When her estranged mother suddenly dies, a wom...
40	The slacker owner of a public bath house ralli...
42	Vignettes of the summer holidays follow vacati...
48	After his girlfriend wins the Miss Nigeria pag...
49	When a carefree bachelor is unexpectedly left ...
72	Nearing a midlife crisis, thirty-something Mik...
77	A bumbling Paris policeman is doggedly determi...
78	Graduate students Krish and Ananya hope to win...
82	Working with an artificial intelligence to inv...

Q10. In how many movies / tv shows “Tom Cruise” was casted ?

```
[24]: len(df[df["Cast"].str.contains("Tom Cruise", case=False, na=False)])
```

```
[24]: 2
```

Q11. What are the different Ratings provided by netflix ?

```
[25]: print(df["Rating"].unique())
```

```
['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']
```

Q12. How many movies got the ‘TV-14’ rating in ‘Canada’ ?

```
[26]: len(df[(df["Rating"]=="TV-14") & (df["Country"]=="Canada")])
```

```
[26]: 23
```

Q13. How many movies get ‘R’ rating in year ‘2019’ ?

```
[27]: len(df[(df["Rating"]=="R") & (df["Release_Date"].dt.year==2019)])
```

```
[27]: 226
```

Q14. What is the max duration of Movie/Show in this dataset ?

```
[28]: df.Duration.value_counts()
```

```
[28]: 1 Season      1608
      2 Seasons    382
      3 Seasons    184
      90 min       136
      93 min       131
      ...
      182 min      1
      224 min      1
      233 min      1
      196 min      1
      191 min      1
      Name: Duration, Length: 216, dtype: int64
```

```
[29]: df[["Minutes","Unit"]] = df["Duration"].str.split(" ", expand=True)
      df[["Minutes","Unit"]].head()
```

```
[29]:   Minutes    Unit
      0      4  Seasons
      1     93     min
      2     78     min
      3     80     min
      4    123     min
```

```
[30]: df["Minutes"].dtype
```

```
[30]: dtype('O')
```

```
[31]: df["Minutes"] = df["Minutes"].astype("int64")
```

```
[32]: df["Minutes"].dtype
```

```
[32]: dtype('int64')
```

```
[33]: print("The max duration in this dataset is ---->>>",df["Minutes"].
      ↪max(),"Minutes")
```

The max duration in this dataset is ---->>> 312 Minutes

Q15. Which Individual Country has maximum number of TV Shows ?

```
[34]: df_tv = df[df["Category"]=="TV Show"]
      df_tv["Country"].value_counts().sort_values(ascending=False).head(1)
      # United States has most number of TV Shows
```

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

Q16. Find out the instances where category is 'movie' and type is 'drama' ?

```
[36]: df[(df["Category"]=="Movie") & (df["Type"].str.contains("Drama", case=False))].
      head(15)
```

```
[36]:   Show_Id Category      Title      Director \
1         s2   Movie      07:19   Jorge Michel Grau
4         s5   Movie         21   Robert Luketic
7         s8   Movie       187   Kevin Reynolds
10        s11   Movie      1922   Zak Hilditch
15        s16   Movie     Oct-01   Kunle Afolayan
17        s18   Movie     22-Jul   Paul Greengrass
18        s19   Movie     15-Aug   Swapnaneel Jayakar
20        s21   Movie Kuch Bheege Alfaaz      Onir
21        s22   Movie      Goli Soda 2   Vijay Milton
22        s23   Movie   Maj Rati Keteki   Santwana Bardoloi
23        s24   Movie      Mayurakshi   Atanu Ghosh
31        s32   Movie   #FriendButMarried   Rako Prijanto
32        s33   Movie   #FriendButMarried 2   Rako Prijanto
36        s37   Movie      #Selfie   Cristina Jacob
37        s38   Movie   #Selfie 69   Cristina Jacob
```

```
Cast \
1   Demián Bichir, Héctor Bonilla, Oscar Serrano, ...
4   Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...
7   Samuel L. Jackson, John Heard, Kelly Rowan, Cl...
10  Thomas Jane, Molly Parker, Dylan Schmid, Kaitl...
15  Sadiq Daba, David Bailie, Kayode Olaiya, Kehin...
17  Anders Danielsen Lie, Jon Øigarden, Jonas Stra...
18  Rahul Pethe, Mrunmayee Deshpande, Adinath Koth...
20  Geetanjali Thapa, Zain Khan Durrani, Shray Rai...
21  Samuthirakani, Bharath Seeni, Vinoth, Esakki B...
22  Adil Hussain, Shakil Imtiaz, Mahendra Rabha, S...
23  Soumitra Chatterjee, Prasenjit Chatterjee, Ind...
31  Adipati Dolken, Vanesha Prescilla, Rendi Jhon,...
32  Adipati Dolken, Mawar de Jongh, Sari Nila, Von...
36  Flavia Hojda, Crina Semciuc, Olimpia Melinte, ...
37  Maia Morgenstern, Olimpia Melinte, Crina Semci...
```

```
Country Release_Date Rating Duration \
1         Mexico  2016-12-23  TV-MA    93 min
4   United States  2020-01-01  PG-13   123 min
7   United States  2019-11-01     R    119 min
10  United States  2017-10-20  TV-MA   103 min
15      Nigeria  2019-09-01  TV-14   149 min
17 Norway, Iceland, United States  2018-10-10     R   144 min
18         India  2019-03-29  TV-14   124 min
20         India  2018-09-01  TV-14   110 min
21         India  2018-09-15  TV-14   128 min
```

22		India	2018-09-15	TV-14	117 min
23		India	2018-09-15	TV-14	100 min
31		Indonesia	2020-05-21	TV-G	102 min
32		Indonesia	2020-06-28	TV-G	104 min
36		Romania	2019-06-01	TV-MA	125 min
37		Romania	2019-06-01	TV-MA	119 min

		Type \
1	Dramas, International Movies	
4	Dramas	
7	Dramas	
10	Dramas, Thrillers	
15	Dramas, International Movies, Thrillers	
17	Dramas, Thrillers	
18	Comedies, Dramas, Independent Movies	
20	Dramas, Independent Movies, International Movies	
21	Action & Adventure, Dramas, International Movies	
22	Dramas, International Movies	
23	Dramas, International Movies	
31	Dramas, International Movies, Romantic Movies	
32	Dramas, International Movies, Romantic Movies	
36	Comedies, Dramas, International Movies	
37	Comedies, Dramas, International Movies	

	Description	Minutes	Unit
1	After a devastating earthquake hits Mexico Cit...	93	min
4	A brilliant group of students become card-coun...	123	min
7	After one of his high school students attacks ...	119	min
10	A farmer pens a confession admitting to his wi...	103	min
15	Against the backdrop of Nigeria's looming inde...	149	min
17	After devastating terror attacks in Norway, a ...	144	min
18	On India's Independence Day, a zany mishap in ...	124	min
20	After accidentally connecting over the Interne...	110	min
21	A taxi driver, a gangster and an athlete strug...	128	min
22	A successful writer returns to the town that l...	117	min
23	When a middle-aged divorcee returns to Kolkata...	100	min
31	Pining for his high school crush for years, a ...	102	min
32	As Ayu and Ditto finally transition from best ...	104	min
36	Two days before their final exams, three teen ...	125	min
37	After a painful breakup, a trio of party-lovin...	119	min

Q17. Find out the instances where category is 'TV Show' and type is 'Kids TV' ?

```
[40]: df[(df["Category"]=="TV Show") & (df["Type"].str.contains("Kids' TV"),
↪case=False)].head(15)
```

```

[40]:
Show_Id Category Title Director \
108 s109 TV Show 3Below: Tales of Arcadia NaN
111 s112 TV Show 44 Cats NaN
225 s226 TV Show A Series of Unfortunate Events NaN
276 s277 TV Show Abby Hatcher Kyran Kelly
364 s365 TV Show Akbar Birbal NaN
380 s381 TV Show Alexa & Katie NaN
396 s397 TV Show Alien TV NaN
411 s412 TV Show All Hail King Julien NaN
412 s413 TV Show All Hail King Julien: Exiled NaN
434 s435 TV Show Alphablocks NaN
523 s524 TV Show Angry Birds NaN
556 s557 TV Show NaN
570 s571 TV Show Archibald's Next Big Thing NaN
598 s599 TV Show Ask the StoryBots NaN
615 s616 TV Show Atomic Puppet NaN

Cast Country \
108 Tatiana Maslany, Diego Luna, Nick Offerman, Ni... United States
111 Sarah Natochenny, Suzy Myers, Simona Berman, E... Italy
225 Neil Patrick Harris, Patrick Warburton, Malina... United States
276 Macy Drouin, Wyatt White, Paul Sun-Hyung Lee, ... United States, Canada
364 Kiku Sharda, Vishal Kotian, Delnaaz Irani India
380 Paris Berelc, Isabel May, Tiffani Thiessen, Em... United States
396 Rob Tinkler, Julie Lemieux, John Cleland, Kyle... NaN
411 Danny Jacobs, Andy Richter, Henry Winkler, Kev... United States
412 Danny Jacobs, Andy Richter, Kevin Michael Rich... NaN
434 Teresa Gallagher, David Holt, Lizzie Waterworth United Kingdom
523 Antti Pääkkönen, Heljä Heikkinen, Lynne Guagli... Finland
556 NaN Japan
570 Tony Hale, Rosamund Pike, Jordan Fisher, Chels... NaN
598 Judy Greer, Erin Fitzgerald, Fred Tatasciore, ... United States
615 Eric Bauza, Lisa Norton, Carlos Díaz, Peter Ol... Canada, France

Release_Date Rating Duration \
108 2019-07-12 TV-Y7 2 Seasons
111 2020-10-01 TV-Y7 2 Seasons
225 2019-01-01 TV-PG 3 Seasons
276 2020-07-01 TV-Y 1 Season
364 2020-03-31 TV-G 1 Season
380 2020-06-13 TV-G 4 Seasons
396 2020-08-21 TV-Y7 1 Season
411 2017-12-01 TV-Y7 5 Seasons
412 2017-05-12 TV-Y7 1 Season
434 2020-05-25 TV-Y 5 Seasons
523 2019-03-16 TV-Y7 3 Seasons
556 2018-12-23 TV-Y7 2 Seasons

```

570	2020-03-20	TV-Y7	2 Seasons
598	2019-08-02	TV-Y	3 Seasons
615	2017-12-01	TV-Y7	1 Season

	Type \
108	Kids' TV, TV Action & Adventure, TV Sci-Fi & F...
111	Kids' TV
225	Kids' TV, TV Action & Adventure, TV Comedies
276	Kids' TV
364	Kids' TV, TV Comedies, TV Dramas
380	Kids' TV, TV Comedies
396	Kids' TV, TV Comedies
411	Kids' TV, TV Comedies
412	Kids' TV, TV Action & Adventure, TV Comedies
434	Kids' TV
523	Kids' TV, TV Comedies
556	Anime Series, Kids' TV
570	Kids' TV, TV Comedies
598	Kids' TV
615	Crime TV Shows, Kids' TV, TV Comedies

	Description	Minutes	Unit
108	After crash-landing on Earth, two royal teen a...	2	Seasons
111	Paw-esome tales abound when singing furry frie...	2	Seasons
225	The extraordinary Baudelaire orphans face tria...	3	Seasons
276	A big-hearted girl helps her Fuzzly friends wh...	1	Season
364	From battles of wit to fights for justice, Emp...	1	Season
380	Alexa is battling cancer. But with her best fr...	4	Seasons
396	Alien reporters Ixbee, Pixbee and Squee travel...	1	Season
411	In this Emmy winner for Outstanding Children's...	5	Seasons
412	Julien's been dethroned, but loyal friends and...	1	Season
434	The letters of the alphabet come to life in Al...	5	Seasons
523	Birds Red, Chuck and their feathered friends h...	3	Seasons
556	Hailing from the mountains of Iga, Kanzo Hatto...	2	Seasons
570	Happy-go-lucky chicken Archibald may not remem...	2	Seasons
598	Five curious little creatures track down the a...	3	Seasons
615	Captain Atomic - once a superhero, now a sock ...	1	Season