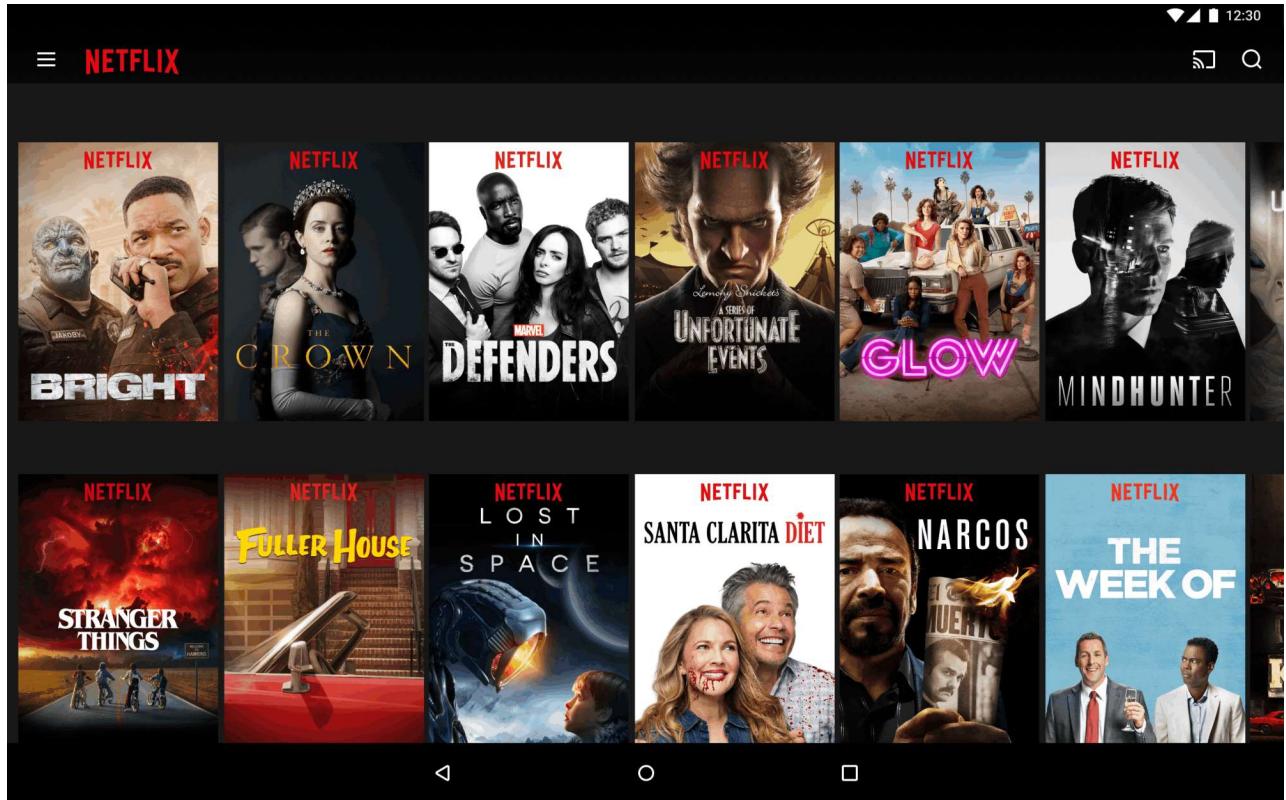


Netflix EDA



- EDA Questions

- What different types of show or movie are uploaded on Netflix?
- What is the Correlation between the features?
- Most watched shows on the Netflix?
- Distribution of Ratings?
- Which has the highest rating Tv show or Movies?
- Finding the best Month for releasing content?
- Highest watched genres on Netflix?
- Released movie over the years?
- movies made on year basis?
- What is the show id and director for 'House of cards'?
- Show all the movies that were released in year 2000.
- Show only the title of all TV shows that were released in India only.
- Show top 10 director, who gave the highest number of TV shows & Movies to Netflix?
- In how many movies/ tv shows, 'tom Cruise' was cast?
- How many movies got the "TV-14" rating in the caneda?

```
In [1]: %%time
import numpy as np          # for numerical computing
import pandas as pd         # for dataframe and manipulation
import seaborn as sns        # for graph and data visualization
from matplotlib import pyplot as plt
import pandas_profiling as prf
sns.set()
%matplotlib inline
```

Wall time: 12.4 s

<timed exec>:5: DeprecationWarning: `import pandas_profiling` is going to be deprecated by April 1st. Please use `import ydata_profiling` instead.

```
In [2]: # Load the data
Neftlix_data=pd.read_csv("F:\INEURON\Python class\EDA-on-NETFLIX-python-project-main/netflix_titles_2021
```

In [3]: `Neftlix_data # read the data`

Out[3]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	des...
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	fath...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	pe...Caj...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	Tch...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	fla...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In c...
...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A ca...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	Wh...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Lo...
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	from...
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A...bo...

8807 rows × 12 columns

- showid: Gives the information about showid
- type: Gives information about 2 different unique values one is TV Show and another is Movie.

- title: Gives information about the title of Movie or TV Show.
- director: Gives information about the director who directed the Movie or TV Show.
- cast: Gives information about the cast who plays role in Movie or TV Show.
- country: Gives information about the Name of country.
- date_added: Gives information about the tv shows or movie released.
- release_year: Gives information about the year when Movie or TV Show was released.
- rating: Gives information about the Movie or TV Show are in which category (eg like the movies are only for students, or adults, etc).
- duration: Gives information about the duration of Movie or TV Show.
- listed_in: Gives information about the genre of Movie or TV Show.
- description: Gives information about the description of Movie or TV Show.

In [4]: # top 5 record
Neftlix_data.head()

Out[4]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Nan	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As his father nears the end of his life, he finds himself
1	s2	TV Show	Blood & Water	Nan	Ama Qamata, Khosi Ngema, Gail Mabalane, Thabani...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	Aft crossroads at party, Cape Town
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Nan	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Action...	To protect his family from powerful drug lord
3	s4	TV Show	Jailbirds New Orleans	Nan	Nan	Nan	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feud, flirtation, and toil talk between old and new
4	s5	TV Show	Kota Factory	Nan	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city coaching center known as train

```
In [5]: # bottom 5 record
Neftlix_data.tail()
```

Out[5]:

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey Jr...	United States	November 20, 2019	2007	R 158 min	Cult Movies, Dramas, Thrillers	A political satire
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7 2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While along sp to youn
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R 88 min	Comedies, Horror Movies	Looki survi world t ov
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG 88 min	Children & Family Movies, Comedies	Dra from ci fc superh
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14 111 min	Dramas, International Movies, Music & Musicals	A sci but boy w his wa



```
In [6]: #check the shape of data
Neftlix_data.shape
```

Out[6]: (8807, 12)

```
In [7]: # check size of data
Neftlix_data.size
```

Out[7]: 105684

```
In [8]: # columns present in data
Neftlix_data.columns
```

```
Out[8]: Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
       dtype='object')
```

```
In [9]: # check data type of columns
Neftlix_data.dtypes
```

```
Out[9]: show_id      object
type        object
title       object
director    object
cast        object
country     object
date_added  object
release_year int64
rating      object
duration    object
listed_in   object
description object
dtype: object
```

```
In [10]: # to show index,data types each column name
```

```
Neftlix_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype  
---  --  
 0   show_id     8807 non-null   object  
 1   type        8807 non-null   object  
 2   title       8807 non-null   object  
 3   director    6173 non-null   object  
 4   cast         7982 non-null   object  
 5   country     7976 non-null   object  
 6   date_added  8797 non-null   object  
 7   release_year 8807 non-null   int64  
 8   rating      8803 non-null   object  
 9   duration    8804 non-null   object  
 10  listed_in   8807 non-null   object  
 11  description 8807 non-null   object  
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
```

```
In [11]: # show the statrical information
```

```
Neftlix_data.describe()
```

```
Out[11]: release_year
```

	release_year
count	8807.000000
mean	2014.180198
std	8.819312
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

```
In [12]: # check the how many unique value present in dataset
```

```
Neftlix_data.nunique()
```

```
Out[12]: show_id      8807
type          2
title         8807
director     4528
cast          7692
country       748
date_added   1767
release_year  74
rating         17
duration      220
listed_in     514
description   8775
dtype: int64
```

Data Preparation (Data Cleaning)

```
In [13]: # to check the null value present in data
miss=Neftlix_data.isnull().sum()
miss
```

```
Out[13]: show_id      0
          type        0
          title       0
          director    2634
          cast         825
          country     831
          date_added  10
          release_year 0
          rating       4
          duration     3
          listed_in    0
          description   0
          dtype: int64
```

```
In [14]: # calculating percentage
miss_new=(Neftlix_data.isnull().sum()/len(Neftlix_data))*100
miss_new
```

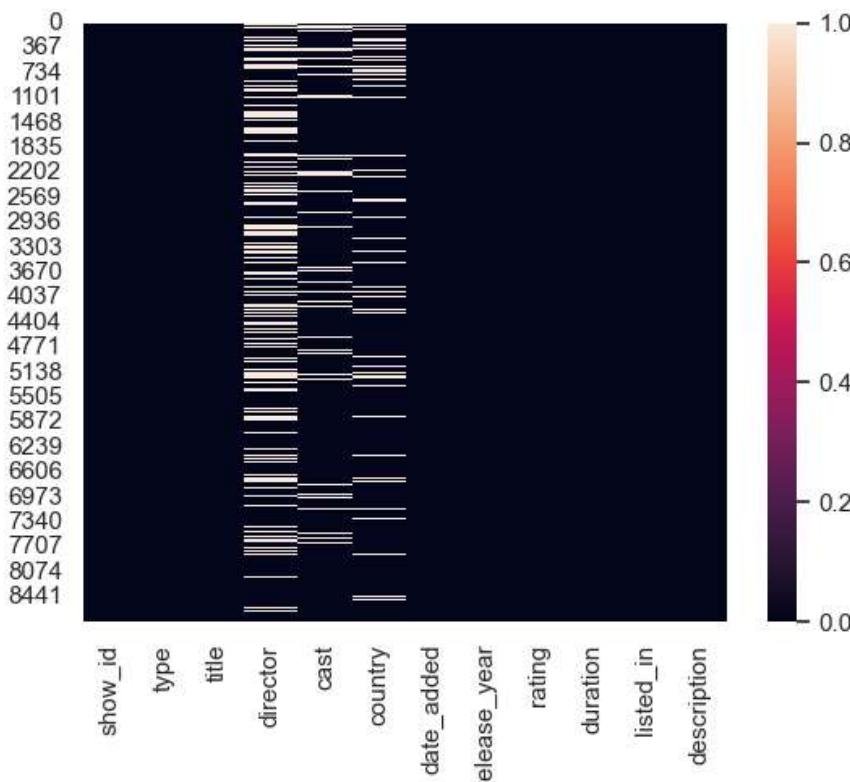
```
Out[14]: show_id      0.000000
          type        0.000000
          title       0.000000
          director    29.908028
          cast         9.367549
          country     9.435676
          date_added  0.113546
          release_year 0.000000
          rating       0.045418
          duration     0.034064
          listed_in    0.000000
          description   0.000000
          dtype: float64
```

```
In [15]: # missing value with percentage
con=pd.concat([miss,miss_new],axis=1,keys=["Total","percent %"])
con
```

	Total	percent %
show_id	0	0.000000
type	0	0.000000
title	0	0.000000
director	2634	29.908028
cast	825	9.367549
country	831	9.435676
date_added	10	0.113546
release_year	0	0.000000
rating	4	0.045418
duration	3	0.034064
listed_in	0	0.000000
description	0	0.000000

```
In [16]: # using heatmap to show the null value count  
sns.heatmap(Neftlix_data.isnull())
```

Out[16]: <AxesSubplot:>



Insights

- We can see that from above graph director,Cast,Country have maximum number of null values.
- We will see how to deal with it
- So we delete director and Cast columns because they are not going to use those features right now

In [5]: # Always take a copy of the original data don't do manipulation on the original data

```
Neftlix_Copy=Neftlix_data.copy()
Neftlix_Copy
```

Out[5]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	des...
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Nan	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	father th...
1	s2	TV Show	Blood & Water	Nan	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	Ca...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Nan	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	Tch...
3	s4	TV Show	Jailbirds New Orleans	Nan	Nan	Nan	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	fla...
4	s5	TV Show	Kota Factory	Nan	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In...
...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	Aca...
8803	s8804	TV Show	Zombie Dumb	Nan	Nan	Nan	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	Wh...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Lo...
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	from...
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A...

8807 rows × 12 columns

```
In [18]: # dropping the na value the director and cast columns
Neftlix_Copy=Neftlix_Copy.dropna(how='any',subset=['director','cast'])
Neftlix_Copy.head(2)
```

Out[18]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	Season 1	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	Season 1	TV Dramas, TV Horror, TV Mysteries	The arrival of a charismatic young priest brin...

```
In [19]: Neftlix_Copy.isnull().sum()
```

Out[19]:

show_id	0
type	0
title	0
director	0
cast	0
country	364
date_added	0
release_year	0
rating	1
duration	3
listed_in	0
description	0
dtype: int64	

Observation

- from above observation we can see that our NA value got removed which present in director and cast columns

```
In [20]: Neftlix_Copy.fillna({'country':'missing','rating':'missing'},inplace=True)
```

C:\Users\em\AppData\Local\Temp\ipykernel_8368\32987932.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
Neftlix_Copy.fillna({'country':'missing','rating':'missing'},inplace=True)
```

```
In [21]: Neftlix_Copy.isnull().sum()
```

Out[21]:

show_id	0
type	0
title	0
director	0
cast	0
country	0
date_added	0
release_year	0
rating	0
duration	3
listed_in	0
description	0
dtype: int64	

Pre Profiling

Now performing data profiling to understand data better

```
In [22]: Neftlix_Profile=prf.ProfileReport(Neftlix_Copy)
Neftlix_Profile
```

```
Summarize dataset: 0% | 0/5 [00:00<?, ?it/s]
Generate report structure: 0% | 0/1 [00:00<?, ?it/s]
Render HTML: 0% | 0/1 [00:00<?, ?it/s]
```

Out[22]:

```
In [23]: Neftlix_Profile.to_file(output_file="netflix before preprocessing.html")
```

```
Export report to file: 0% | 0/1 [00:00<?, ?it/s]
```

Preprocessing

- Dealing with duplicate rows
 - Find number of duplicate rows in dataset
 - Print duplicate entries and analyse
 - Drop duplicate entries from dataset

```
In [24]: # to check the duplicate rows present in dataset
Neftlix_Copy[Neftlix_Copy.duplicated()]
```

```
Out[24]: show_id  type  title  director  cast  country  date_added  release_year  rating  duration  listed_in  description
```

```
In [25]: # to count the duplicated rows
Neftlix_Copy.duplicated().sum()
```

Out[25]: 0

Observation

- No duplicate rows present in our dataset

EDA Questions

What different types of show or movie are uploaded on Netflix? or

How many TV shows & Movies are in the dataset? Show with Bar Graph?

```
In [26]: Neftlix_Copy.columns
```

```
Out[26]: Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
       dtype='object')
```

In [27]: `Neftlix_Copy.head(2)`

		show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
2		s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	September 24, 2021	2021	TV-MA	Season 1	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
5		s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	September 24, 2021	2021	TV-MA	Season 1	TV Dramas, TV Horror, TV Mysteries	The arrival of a charismatic young priest brin...

In [28]: `#method-1`

```
Neftlix_Copy.groupby("type")["title"].count().sort_values(ascending=True)
```

Out[28]: type

```
TV Show    178
```

```
Movie     5522
```

```
Name: title, dtype: int64
```

In [29]: `#method-2`

```
Neftlix_Copy.type.value_counts().to_frame('value_count')
```

Out[29]:

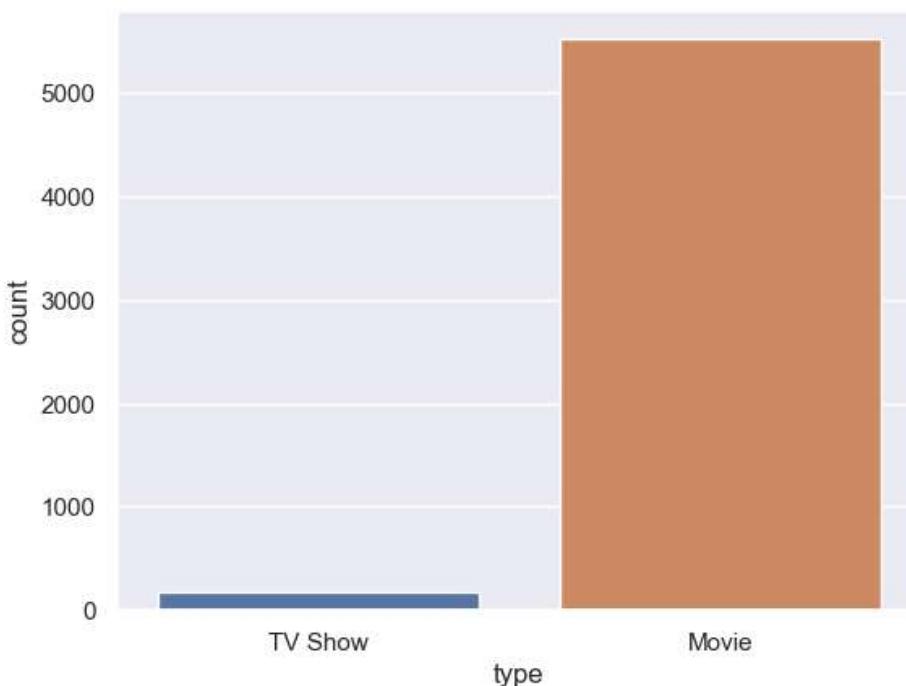
	value_count
Movie	5522
TV Show	178

In [30]: `# to show the graph between movie vs TV show`

```
sns.countplot(Neftlix_Copy['type'])
```

```
C:\Users\em\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(
```

Out[30]: <AxesSubplot:xlabel='type', ylabel='count'>



Observation

- There are 5522 types of movies are uploaded
- and 178 types of TV shows are uploaded

What is the Correlation between the features?

In [31]: Neftlix_Copy.dtypes

```
Out[31]: show_id      object
type          object
title         object
director      object
cast          object
country        object
date_added    object
release_year   int64
rating         object
duration       object
listed_in      object
description    object
dtype: object
```

- Converting object into datetime formate of 'date_added'

In [32]: # converting date_added obbject type to datetime format

```
Neftlix_Copy['date_added']=pd.to_datetime(Neftlix_Copy["date_added"])
Neftlix_Copy.head(2)
```

C:\Users\em\AppData\Local\Temp\ipykernel_8368\2363749807.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
Neftlix_Copy['date_added']=pd.to_datetime(Neftlix_Copy["date_added"])
```

Out[32]:

		show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
2	s3	TV Show	Ganglands	Julien Ledercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1	Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1	Season	TV Dramas, TV Horror, TV Mysteries	The arrival of a charismatic young priest brin...

- Now ,we create two feature year_added and month_added from date_added column

```
In [33]: # to add new columns year_added and month_added in our dataset from extracting
# to date_added
Neftlix_Copy["year_added"] = Neftlix_Copy["date_added"].dt.year
Neftlix_Copy["month_added"] = Neftlix_Copy["date_added"].dt.month
Neftlix_Copy.head(2)
```

C:\Users\em\AppData\Local\Temp\ipykernel_8368\1202751847.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

Neftlix_Copy["year_added"] = Neftlix_Copy["date_added"].dt.year

C:\Users\em\AppData\Local\Temp\ipykernel_8368\1202751847.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

Neftlix_Copy["month_added"] = Neftlix_Copy["date_added"].dt.month

		show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
2	s3	TV Show	Ganglands	Julien Ledercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	Season 1	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...	
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	Season 1	TV Dramas, TV Horror, TV Mysteries	The arrival of a charismatic young priest brin...	

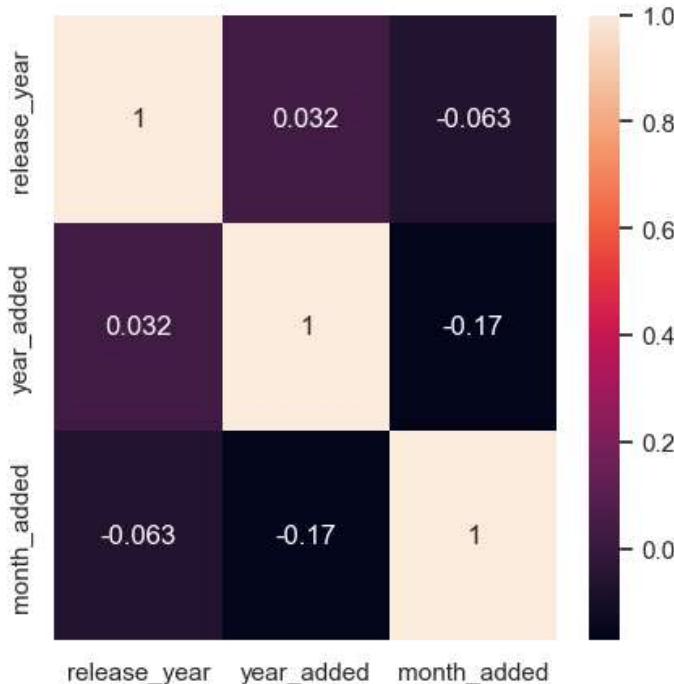
```
In [34]: Neftlix_Copy.dtypes
```

```
Out[34]: show_id          object
type            object
title           object
director        object
cast            object
country          object
date_added      datetime64[ns]
release_year    int64
rating          object
duration        object
listed_in       object
description     object
year_added      int64
month_added     int64
dtype: object
```

```
In [35]: # calculating the corelation between the release_year,year_added,month_added
Neftlix_Copy.corr()
```

	release_year	year_added	month_added
release_year	1.000000	0.032265	-0.062903
year_added	0.032265	1.000000	-0.170078
month_added	-0.062903	-0.170078	1.000000

```
In [36]: # show graph of coorelation between release_year,year_added,month_added
plt.subplots(figsize=(5,5))
sns.heatmap(Neftlix_Copy.corr(), annot=True)
plt.show()
```



Observation

- coorelation coefficient ranges between 1 to -1
- coorelation 1 means strong positive coorelation
- coorelation 0 means no coorelation
- coorelation -1 means strong negative coorelation

Most watched shows on the Netflix?

```
In [37]: Neftlix_Copy.head(2)
```

		show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
2	s3	TV Show	Ganglands	Julien Ledercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...	
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	The arrival of a charismatic young priest brin...	

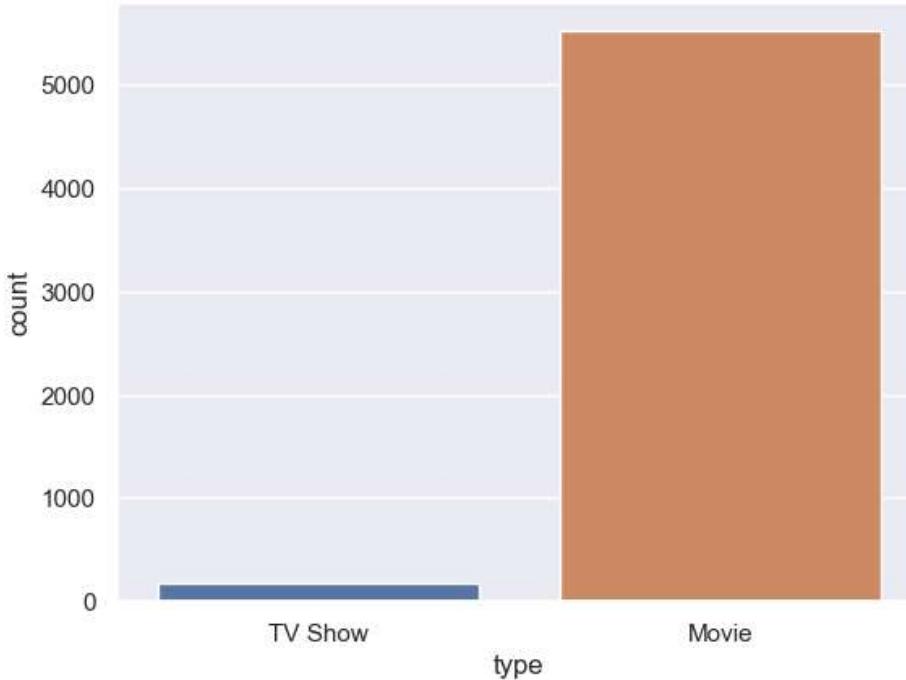
```
In [38]: Neftlix_Copy.groupby('type')['title'].count().sort_values(ascending=True)
```

```
Out[38]: type
TV Show      178
Movie       5522
Name: title, dtype: int64
```

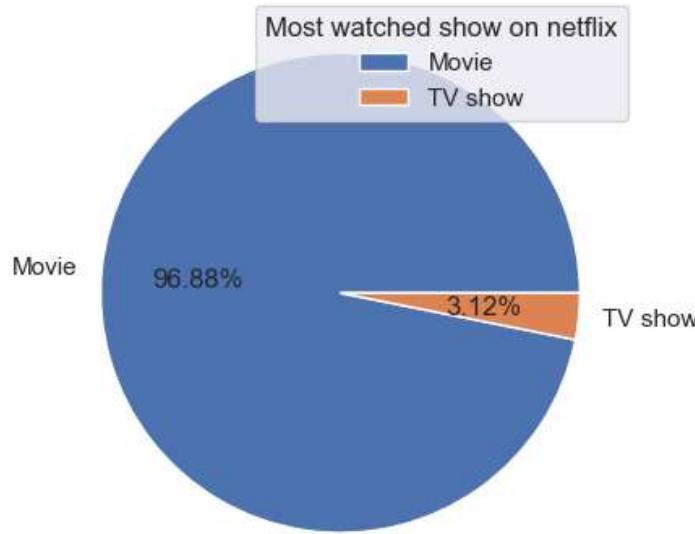
```
In [39]: sns.countplot(Neftlix_Copy['type'])
```

C:\Users\em\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(

```
Out[39]: <AxesSubplot:xlabel='type', ylabel='count'>
```



```
In [40]: #by using pie chart
type_show=['Movie', "TV show"]
value_count=[5522,178]
plt.pie(value_count,labels=type_show,autopct="%2.2f%%")
plt.legend(title="Most watched show on netflix")
plt.show()
```



Observation

- mostly people watch movies(96%) rather than TV movies(3.12%)

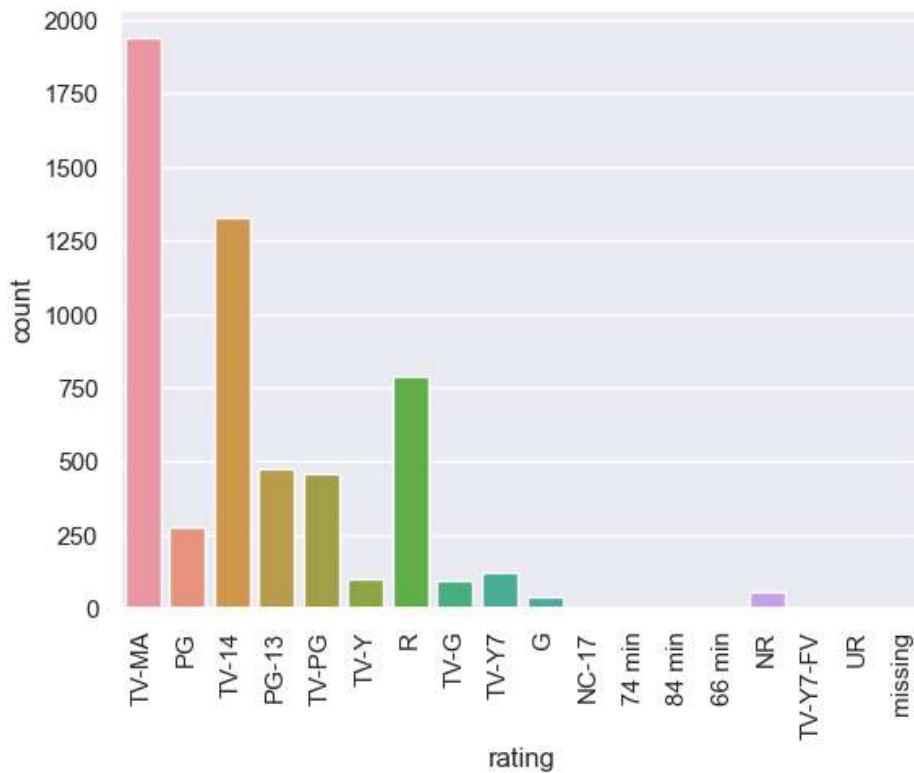
Distribution of Ratings? & and what are different rating defined by Netflix

```
In [41]: Neftlix_Copy.head(2)
```

		show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...	
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	The arrival of a charismatic young priest brin...	

```
In [42]: # countplot for distribution
sns.countplot(x=Neftlix_Copy['rating'])
plt.xticks(rotation=90)
```

```
Out[42]: (array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
       17]),  
 [Text(0, 0, 'TV-MA'),  
  Text(1, 0, 'PG'),  
  Text(2, 0, 'TV-14'),  
  Text(3, 0, 'PG-13'),  
  Text(4, 0, 'TV-PG'),  
  Text(5, 0, 'TV-Y'),  
  Text(6, 0, 'R'),  
  Text(7, 0, 'TV-G'),  
  Text(8, 0, 'TV-Y7'),  
  Text(9, 0, 'G'),  
  Text(10, 0, 'NC-17'),  
  Text(11, 0, '74 min'),  
  Text(12, 0, '84 min'),  
  Text(13, 0, '66 min'),  
  Text(14, 0, 'NR'),  
  Text(15, 0, 'TV-Y7-FV'),  
  Text(16, 0, 'UR'),  
  Text(17, 0, 'missing')])
```



```
In [43]: Neftlix_Copy['rating'].nunique()
```

```
Out[43]: 18
```

Observation

- most people like TV-MA and TV-14 and less prefer NC-17
- There are 18 types of rating in netflix

Which has the highest rating Tv show or Movies?

```
In [44]: Neftlix_Copy.groupby('type')[['rating']].agg(pd.Series.mode)
```

```
Out[44]: type  
Movie      TV-MA  
TV Show    TV-MA  
Name: rating, dtype: object
```

Adding New feature

- new genre(Genre 1,Genre 2,Genre 3)

```
In [45]: new_genre=Neftlix_Copy['listed_in'].str.split(", ",2)  
new_genre
```

```
Out[45]: 2      [Crime TV Shows, International TV Shows, TV ...  
5          [TV Dramas, TV Horror, TV Mysteries]  
6          [Children & Family Movies]  
7      [Dramas, Independent Movies, International M...  
8          [British TV Shows, Reality TV]  
...  
8801      [Dramas, International Movies, Thrillers]  
8802      [Cult Movies, Dramas, Thrillers]  
8804      [Comedies, Horror Movies]  
8805      [Children & Family Movies, Comedies]  
8806      [Dramas, International Movies, Music & Music...  
Name: listed_in, Length: 5700, dtype: object
```

```
In [46]: #The str.get(0) function is applied to the 'new_genre' column, which contains the lists of split values.
# This function retrieves the element at index 0 from each list and assigns it to the 'Genre 1' column
Neftlix_Copy['Genre 1']=new_genre.str.get(0)
Neftlix_Copy['Genre 2']=new_genre.str.get(1)
Neftlix_Copy['Genre 3']=new_genre.str.get(2)
Neftlix_Copy.head(2)
```

C:\Users\em\AppData\Local\Temp\ipykernel_8368\2040198132.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
Neftlix_Copy['Genre 1']=new_genre.str.get(0)
C:\Users\em\AppData\Local\Temp\ipykernel_8368\2040198132.py:4: SettingWithCopyWarning:  

A value is trying to be set on a copy of a slice from a DataFrame.  

Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
Neftlix_Copy['Genre 2']=new_genre.str.get(1)
C:\Users\em\AppData\Local\Temp\ipykernel_8368\2040198132.py:5: SettingWithCopyWarning:  

A value is trying to be set on a copy of a slice from a DataFrame.  

Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
Neftlix_Copy['Genre 3']=new_genre.str.get(2)
```

Out[46]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	Season 1	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	Season 1	TV Dramas, TV Horror, TV Mysteries	The arrival of a charismatic young priest brin...

In [47]: Neftlix_Copy['Genre 1'].describe(include=all)

```
Out[47]: count      5700
unique     33
top      Dramas
freq      1586
Name: Genre 1, dtype: object
```

In [48]: Neftlix_Copy['Genre 2'].describe(include=all)

```
Out[48]: count      4485
unique     35
top      International Movies
freq      1545
Name: Genre 2, dtype: object
```

```
In [49]: Neftlix_Copy['Genre 3'].describe(include=all)
```

```
Out[49]: count          2434
unique           29
top    International Movies
freq            881
Name: Genre 3, dtype: object
```

Now drop the column listed_in because no use of that

```
In [50]: Neftlix_Copy.drop('listed_in',axis=1,inplace=True)
Neftlix_Copy.head(2)
```

C:\Users\em\AppData\Local\Temp\ipykernel_8368\57628386.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
Neftlix_Copy.drop('listed_in',axis=1,inplace=True)
```

```
Out[50]:
```

		show_id	type	title	director	cast	country	date_added	release_year	rating	duration	description	year_added
2	s3	TV Show	Ganglands	Julien Ledercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	To protect his family from a powerful drug lor...	2021	
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1 Season	The arrival of a charismatic young priest brin...	2021	

Finding the best Month for releasing content?

```
In [51]: # converting month number to month name
Neftlix_Copy['month_final']=Neftlix_Copy['month_added'].replace({1:'Jan',2:'feb',3:'Mar',4:'Apr',5:'May'
Neftlix_Copy.head(2)
```

C:\Users\em\AppData\Local\Temp\ipykernel_8368\359372360.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
Neftlix_Copy['month_final']=Neftlix_Copy['month_added'].replace({1:'Jan',2:'feb',3:'Mar',4:'Apr',5:'May',6:'June',7:'July',8:'Aug',9:'Sept',10:'Oct',11:'Nov',12:'Dec'})
```

Out[51]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	description	year_added
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	To protect his family from a powerful drug lord...	2021
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1 Season	The arrival of a charismatic young priest brings...	2021

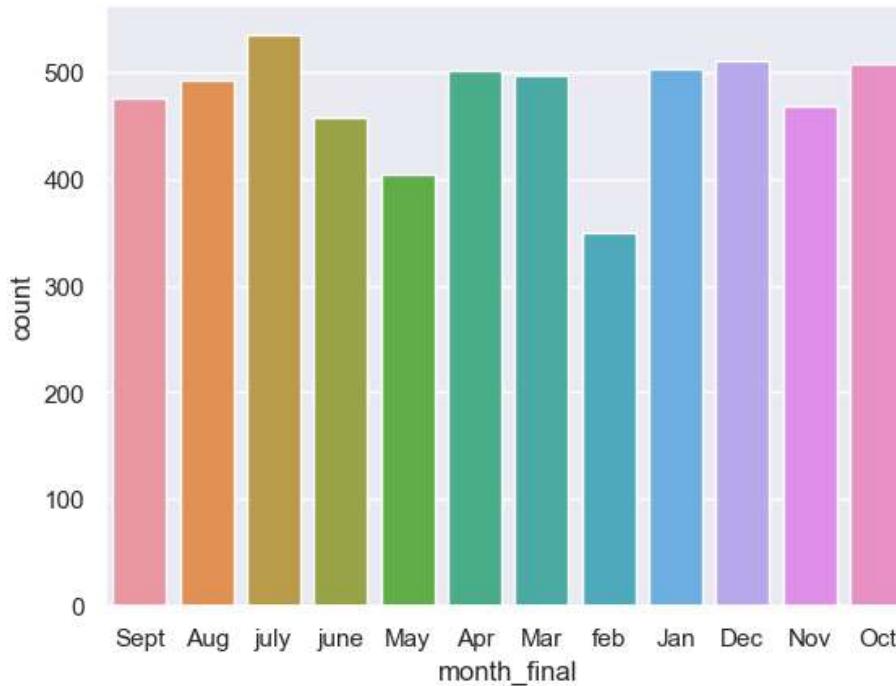
In [52]: Neftlix_Copy.month_final.value_counts().to_frame('value_counts')

Out[52]:

	value_counts
july	535
Dec	511
Oct	507
Jan	503
Apr	501
Mar	497
Aug	492
Sept	475
Nov	468
June	457
May	405
feb	349

```
In [53]: sns.countplot(Neftlix_Copy['month_final'])
plt.show()
```

C:\Users\em\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(



Observation

- We can see that July is the best month has maximum number of movie released
- july is the best month

Highest watched genres on Netflix?

```
In [54]: Neftlix_Copy['Genre 1'].describe(include=all)
```

```
Out[54]: count      5700
unique       33
top        Dramas
freq       1586
Name: Genre 1, dtype: object
```

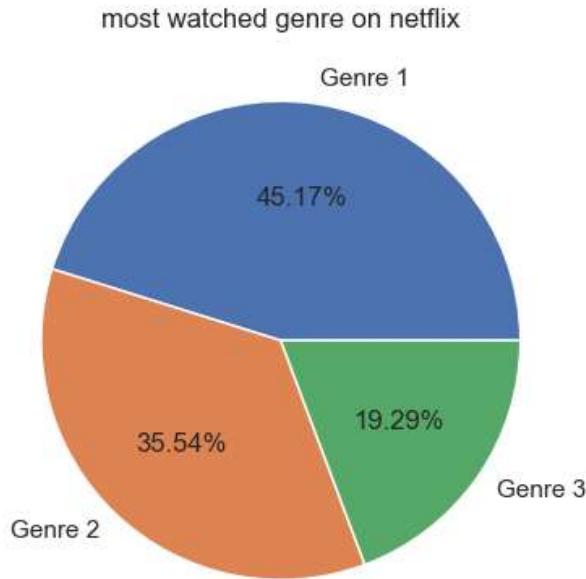
```
In [55]: Neftlix_Copy['Genre 2'].describe(include=all)
```

```
Out[55]: count      4485
unique       35
top    International Movies
freq       1545
Name: Genre 2, dtype: object
```

```
In [56]: Neftlix_Copy['Genre 3'].describe(include=all)
```

```
Out[56]: count      2434
unique       29
top    International Movies
freq       881
Name: Genre 3, dtype: object
```

```
In [57]: # by using pie chart
Genre_type=['Genre 1','Genre 2','Genre 3']
value_count=[5700,4485,2434]
plt.pie(value_count,labels=Genre_type,autopct="%2.2f%%")
plt.title("most watched genre on netflix")
plt.show()
```



Observation

- above pie chart Genre 1 watch maximum times

Released movie over the years?

```
In [58]: Neftlix_Copy.head(2)
```

Out[58]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	description	year_added
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	To protect his family from a powerful drug lor...	2021
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1 Season	The arrival of a charismatic young priest brin...	2021

```
In [59]: Neftlix_Copy.groupby(['release_year'])['release_year'].count().sort_values(ascending=False)
```

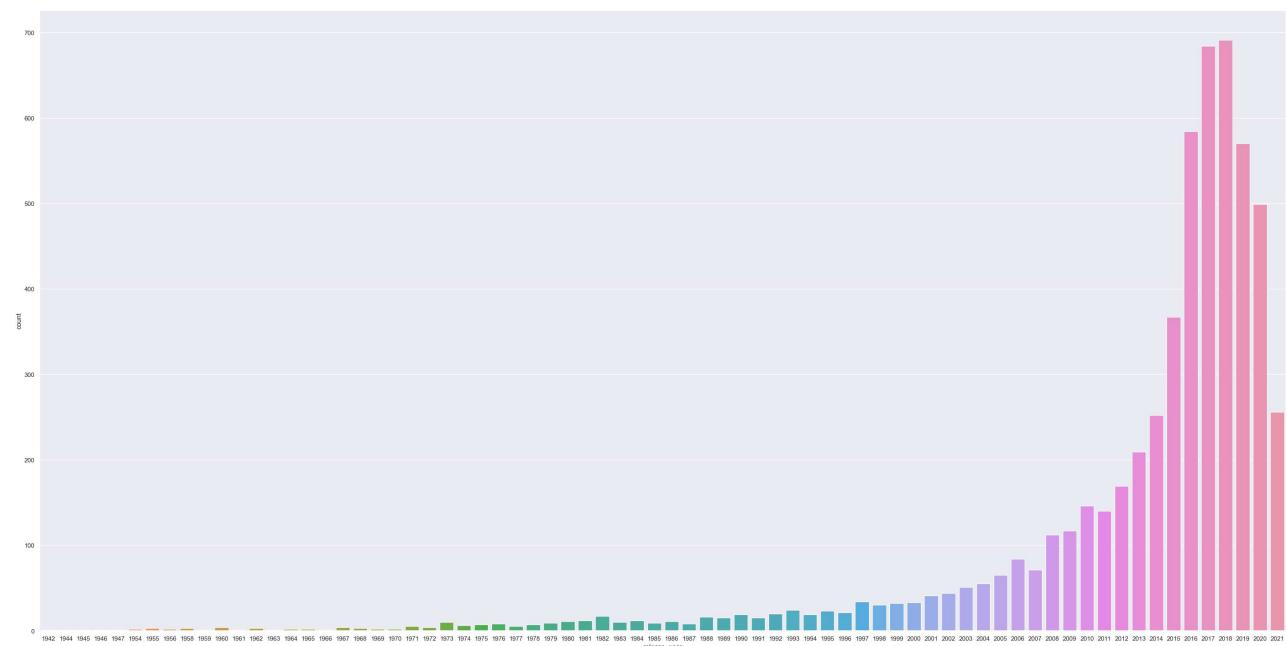
```
Out[59]: release_year
2018    691
2017    684
2016    584
2019    570
2020    499
...
1944     1
1947     1
1946     1
1945     1
1942     1
Name: release_year, Length: 72, dtype: int64
```

```
In [60]: Neftlix_Copy.release_year.value_counts().to_frame('value_count')
```

```
Out[60]:   value_count
2018        691
2017        684
2016        584
2019        570
2020        499
...
...
1946         1
1959         1
1942         1
1947         1
1944         1
```

72 rows × 1 columns

```
In [62]: sns.countplot(x='release_year', data=Neftlix_Copy)
sns.set(rc={'figure.figsize':(40,20)})
plt.show()
```



Observation

- from above graph we can see that sudden increase in creating of movies is from 2015 to 2021

movies made on year basis? or

In which year highest number of the TV shows & Movies were released? show with Bar Graph.

In [63]: Neftlix_Copy.head(2)

Out[63]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	description	year_added
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	To protect his family from a powerful drug lor...	2021
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1 Season	The arrival of a charismatic young priest brin...	2021

In [66]: Neftlix_Copy.groupby('year_added')[['type']].value_counts().sort_values(ascending=False)

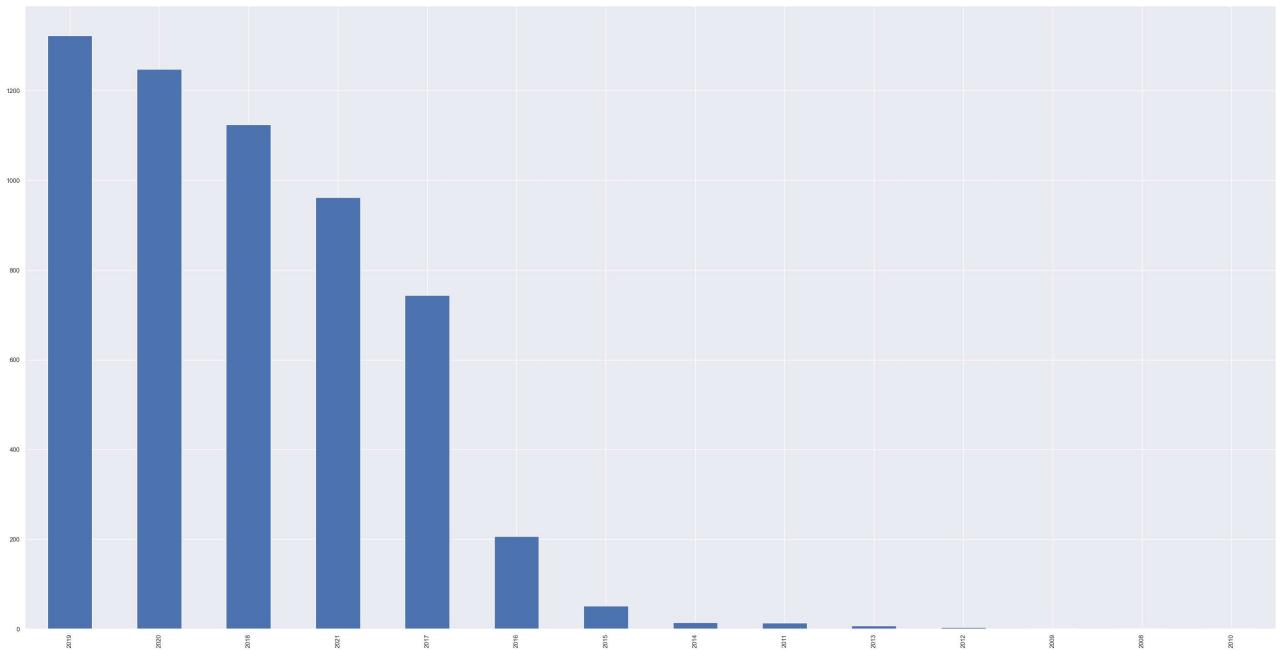
Out[66]:

year_added	type	value
2019	Movie	1293
2020	Movie	1202
2018	Movie	1107
2021	Movie	915
2017	Movie	718
2016	Movie	199
2015	Movie	48
2021	TV Show	47
2020	TV Show	46
2019	TV Show	30
2017	TV Show	26
2018	TV Show	18
2014	Movie	14
2011	Movie	13
2016	TV Show	7
2013	Movie	6
2015	TV Show	3
2012	Movie	3
2009	Movie	2
2013	TV Show	1
2010	Movie	1
2008	Movie	1

Name: type, dtype: int64

In [67]: Neftlix_Copy['year_added'].value_counts().plot(kind='bar')

Out[67]: <AxesSubplot:>



Observation

- We can clearly see that maximum number of movies made on 2019

Show all the movies that were released in year 2000.

In [68]: Neftlix_Copy.head(2)

Out[68]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	description	year_added
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	To protect his family from a powerful drug lord...	2021
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1 Season	The arrival of a charismatic young priest brings...	2021

```
In [82]: # to fetech a record type is movie and release year is 2000
Neftlix_Copy[(Neftlix_Copy['type']=='Movie') & (Neftlix_Copy['release_year']==2000)]
```

3472	s3473	Movie	Rugrats in Paris: The Movie	Stig Bergqvist, Paul Demeyer	Elizabeth Daily, Tara Strong, Cheryl Chase, Ch...	Germany, United States	2019-10-01	2000	G	79 min	The Rugrats take to the big screen and visit P...			
4545	s4546	Movie	Monty Python: Before the Flying Circus	Will Yapp	Graham Chapman, Eric Idle, Terry Jones, Michael...	United Kingdom	2018-10-02	2000	TV-MA	56 min	Discover how six seemingly ordinary bu...			
4723	s4724	Movie	Fiza	Khalid Mohamed	Karisma Kapoor, Jaya Bhaduri, Hrithik Roshan, ...	India	2018-08-02	2000	TV-14	163 min	Fiza's brother disappears during Mumbai's horr...			

```
In [72]:
```

```
Out[72]:
```

```
show_id  type  title  director  cast  country  date_added  release_year  rating  duration  description  year_added  month_added
```

```
In [83]: # to check the shape of data
```

```
Neftlix_Copy[(Neftlix_Copy['type']=='Movie')&(Neftlix_Copy['release_year']==2000)].shape
```

```
Out[83]: (33, 17)
```

Observation

- There are total 33 movies released in the year 2000

Show only the title of all TV shows that were released in India only.

```
In [92]: Neftlix_Copy[(Neftlix_Copy['type']=='TV Show') & (Neftlix_Copy['country']=='India')]['title']
```

```
Out[92]: 297                               Navarasa
1293                               Pitta Kathalu
1520                               Paava Kadhaigal
3036                Jamtara - Sabka Number Ayega
4685                               GHOUL
6024                               7 (Seven)
8172    Th Eena Meena Deeka Chase Comedy Show
Name: title, dtype: object
```

```
In [90]: Neftlix_Copy[(Neftlix_Copy['type']=='TV Show') & (Neftlix_Copy['country']=='India')].shape
```

```
Out[90]: (7, 17)
```

Show top 10 director, who gave the highest number of TV shows & Movies to Netflix?

In [94]: `Neftlix_Copy.head(2)`

Out[94]:

		show_id	type	title	director	cast	country	date_added	release_year	rating	duration	description	year_added
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	To protect his family from a powerful drug lord...	2021	
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1 Season	The arrival of a charismatic young priest brings...	2021	

In [96]: `Neftlix_Copy['director'].value_counts().head(10)`

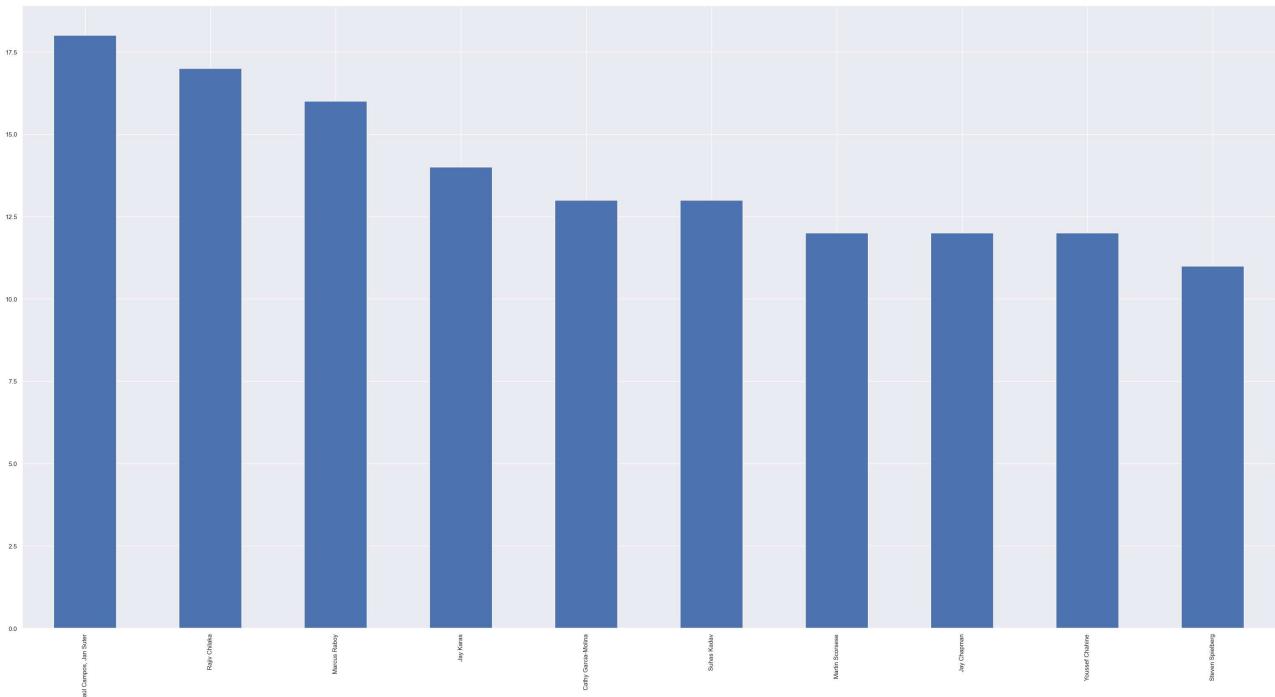
Out[96]:

Raúl Campos, Jan Suter	18
Rajiv Chilaka	17
Marcus Raboy	16
Jay Karas	14
Cathy Garcia-Molina	13
Suhas Kadav	13
Martin Scorsese	12
Jay Chapman	12
Youssef Chahine	12
Steven Spielberg	11

Name: director, dtype: int64

In [97]: `Neftlix_Copy['director'].value_counts().head(10).plot(kind='bar')`

Out[97]: <AxesSubplot:>



Observation

- Raúl Campos, Jan Suter release the maximum number of TV show in india

In how many movies/ tv shows, 'tom Cruise' was cast?

In [98]: Neftlix_Copy.head(2)

Out[98]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	description	year_added
2	s3	TV Show	Ganglands	Julien Ledercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	To protect his family from a powerful drug lor...	2021
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1 Season	The arrival of a charismatic young priest brin...	2021

In [109]: Neftlix_Copy[Neftlix_Copy['cast']=='Tom Cruise']

Out[109]:

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	description	year_added	month_added
2	s3	TV Show	Ganglands	Julien Ledercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	To protect his family from a powerful drug lor...	2021

Nan Value are present first fill null values

In [110]: Neftlix_Copy.dropna()
Neftlix_Copy.head(2)

Out[110]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	description	year_added
2	s3	TV Show	Ganglands	Julien Ledercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	To protect his family from a powerful drug lor...	2021
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1 Season	The arrival of a charismatic young priest brin...	2021

In [112]: Neftlix_Copy[Neftlix_Copy['cast'].str.contains('Tom Cruise')]

Out[112]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	description	year_added
341	s342	Movie	Magnolia	Paul Thomas Anderson	John C. Reilly, Philip Baker Hall, Tom Cruise,...	United States	2021-08-01	1999	R	189 min	Through chance, history and divine intervention...	202
1254	s1255	Movie	Rain Man	Barry Levinson	Dustin Hoffman, Tom Cruise, Valeria Golino, Ge...	United States	2021-03-01	1988	R	134 min	Motivated by money, a selfish workaholic seeki...	202

Observation

- There are only 2 movies casted by Tom Cruise- Magnolia , Rain Man

How many movies got the "TV-14" rating in the caneda?

In [118]: Neftlix_Copy.head(2)

Out[118]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	description	year_added
2	s3	TV Show	Ganglands	Julien Ledercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	missing	2021-09-24	2021	TV-MA	1 Season	To protect his family from a powerful drug lor...	2021
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	missing	2021-09-24	2021	TV-MA	1 Season	The arrival of a charismatic young priest brin...	2021

In [123]: Neftlix_data[(Neftlix_data['type']=='Movie') & (Neftlix_data['rating']=='TV-14')].shape

Out[123]: (1427, 12)

```
In [125]: Neftlix_data[(Neftlix_data['type']=='Movie') & (Neftlix_data['rating']=='TV-14') & (Neftlix_data['count']>=100)]
```

Out[125]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed
157	s158	Movie	Level 16	Danishka Esterhazy	Katie Douglas, Celina Martin, Peter Outerbridge...	Canada	September 1, 2021	2018	TV-14	102 min	Sci-Fant Thr
1279	s1280	Movie	Made You Look: A True Story About Fake Art	Barry Avrich	NaN	Canada	February 23, 2021	2020	TV-14	90 min	Documenta Internati Mo
2769	s2770	Movie	There's Something in the Water	Ellen Page, Ian Daniel	Ellen Page	Canada	March 27, 2020	2019	TV-14	72 min	Documenta Internati Mo
3424	s3425	Movie	Martin Matte: La vie, la mort...eh la la..!	Jean-François Blais	Martin Matte	Canada	October 14, 2019	2019	TV-14	62 min	Stanc Cor
5972	s5973	Movie	#cats_the_mewvie	Michael Margolis	NaN	Canada	February 5, 2020	2020	TV-14	90 min	Documenta Internati Mo
5973	s5974	Movie	#Roxy	Michael Kennedy	Jake Short, Sarah Fisher, Booboo Stewart, Dann...	Canada	April 10, 2019	2018	TV-14	105 min	Comec Roma Mo
6192	s6193	Movie	At First Light	Jason Stone	Stefanie Scott, Theodore Pellerin, Percy Hynes...	Canada	December 27, 2018	2018	TV-14	91 min	Dramas, S & Fan
6537	s6538	Movie	Counterfeiting in Suburbia	Jason Bourque	Sarah Butler, Larissa Albuquerque, Kayla Walla...	Canada	July 1, 2018	2018	TV-14	88 min	Drar Thri
7334	s7335	Movie	Long Time Running	Jennifer Baichwal, Nicholas de Pencier	Gord Downie, Paul Langlois, Gord Sinclair, Rob...	Canada	November 12, 2017	2017	TV-14	96 min	Documenta Internati Movies, M &
7916	s7917	Movie	Sadie's Last Days on Earth	Michael Seater	Morgan Taylor Campbell, Clark Backo, Ricardo H...	Canada	August 22, 2017	2016	TV-14	94 min	Comedies, Fi & Fan
7936	s7937	Movie	Sardaar ji	Rohit Jugraj	Diljit Dosanjh, Neeru Bajwa, Mandy Takhar, Jas...	Canada	December 1, 2017	2015	TV-14	140 min	Comec Internati Movies, M & Musi
8178	s8179	Movie	The Accountant of Auschwitz	Matthew Shoychet	NaN	Canada	July 1, 2019	2018	TV-14	78 min	Documenta Internati Mo
8566	s8567	Movie	The Wrong Babysitter	George Mendeluk	Daphne Zuniga, Seann Gallagher, Lisa Marie Car...	Canada	June 18, 2019	2017	TV-14	89 min	Thri



```
In [126]: Neftlix_data[(Neftlix_data['type']=='Movie') & (Neftlix_data['rating']=='TV-14') & (Neftlix_data['count']>10)]
Out[126]: (13, 12)
```

Observation

- There are 13 movies got TV-14 rating in Canada

```
In [6]: Netflix_clean=prf.ProfileReport(Neftlix_Copy)
Netflix_clean.to_file(output_file="Netflix_post_preprocessing.html")
```

Summarize dataset: 100%	23/23 [00:04<00:00, 4.62it/s, Completed]
Generate report structure: 100%	1/1 [00:03<00:00, 3.72s/it]
Render HTML: 100%	1/1 [00:01<00:00, 1.42s/it]
Export report to file: 100%	1/1 [00:00<00:00, 21.30it/s]

Conclusion

- We explore the Netflix dataset and saw how to clean the data and then jump into how to visualize the data with
- Exploratory Data Analysis
- We saw some basic and advanced level charts of seaborn and matplotlib like Heatmap, Pie-chart, Bar chart, Countplot. And also visualize a data with
 - Pandas_Profiling :Preprofiling &Post_profiling

Connection to Python to Snowflake

```
In [7]: import snowflake.connector
from sqlalchemy import create_engine
import pandas as pd
```

```
In [8]: account = 'ZIJBOYM-LF27852'
user = 'SNOW'
password = 'Sanjeev5'
database = 'netflix'
schema = 'PUBLIC'
warehouse = 'COMPUTE_WH'
```

```
In [9]: from sqlalchemy import create_engine
```

```
In [10]: connection_string = f"snowflake://{user}:{password}@{account}/{database}/{schema}?warehouse={warehouse}"
```

```
In [11]: engine = create_engine(connection_string)
```

```
In [12]: Neftlix_Copy.to_sql(name="Netflix_Data",con = engine,if_exists="replace",index=False)
```

```
Out[12]: 8807
```

```
In [ ]:
```

