

```
#!/pip install jovian --upgrade --quiet
```

## importing Important Library

```
import jovian
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

## Loading the dataset

```
Netflix_df= pd.read_csv('/home/miracle/Documents/Data_Analysis/netflix_titles.csv')
```

```
Netflix_df.head(5)
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listec
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documenta
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	Internatio TV Shows, Dramas, Myster
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime Sho Internatio TV Shows, A
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuser Reality
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	Internatio TV Sho Romantic Shows, T

```
print(Netflix_df.size, Netflix_df.shape, Netflix_df.ndim)
```

# Data preparation and cleaning

```
Netflix_df.info() #Viewing data summary
```

```
<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
```

```
Netflix_df.nunique()
```

```
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
```

```
Netflix_df.nunique().sum()
```

# Checking and handling null values

```
Netflix_df.isnull().any()
```

```
show_id      False
type         False
title        False
director     False
cast         False
country      False
date_added   True
release_year False
rating       True
duration     True
listed_in    False
description  False
dtype: bool
```

```
#Check if there is any null values
Netflix_df.isnull().values.any()
```

True

```
# Let's display the null values
Netflix_df.isnull().sum()
```

```
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
```

```
total_nullvalues= Netflix_df.isnull().sum().sum()
total_nullvalues
```

4307

```
# percentages of nullvalues
total_cell= Netflix_df.size
```

```
percentage_null= (total_nullvalues/total_cell)*100
percentage_null
```

4.075356723818175

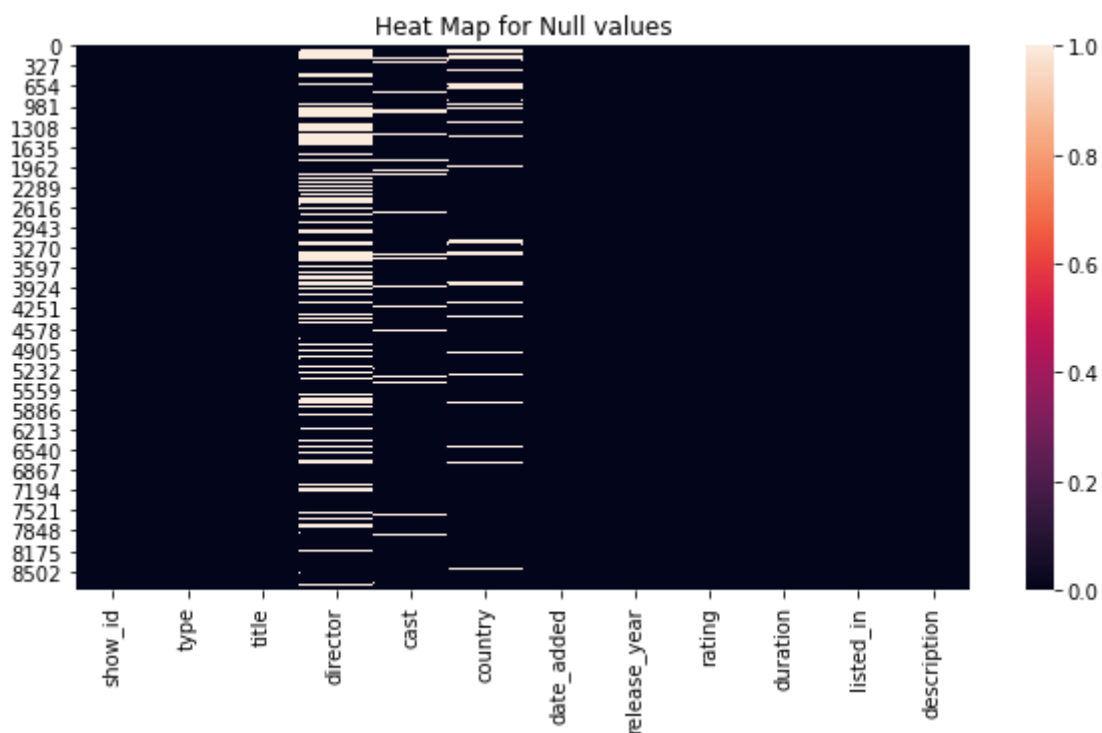
```
Netflix_df.isnull()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	False	False	False	False	True	False	False	False	False	False	False	Fals
1	False	False	False	True	False	False	False	False	False	False	False	Fals
2	False	False	False	False	False	True	False	False	False	False	False	Fals
3	False	False	False	True	True	True	False	False	False	False	False	Fals
4	False	False	False	True	False	False	False	False	False	False	False	Fals
...	...	...	...	...	...	...	...	...	...	...	...	.
8802	False	False	False	False	False	False	False	False	False	False	False	Fals
8803	False	False	False	True	True	True	False	False	False	False	False	Fals
8804	False	False	False	False	False	False	False	False	False	False	False	Fals
8805	False	False	False	False	False	False	False	False	False	False	False	Fals
8806	False	False	False	False	False	False	False	False	False	False	False	Fals

8807 rows × 12 columns

```
# plotting heat map of Netflix_df.isnull()
plt.figure(figsize=(10,5))
sns.heatmap(Netflix_df.isnull())
plt.title('Heat Map for Null values')
```

```
Text(0.5, 1.0, 'Heat Map for Null values')
```



```
# The two ways to deal with Not a Number(NaN) is by (a) fillna and (b) dropna.
#We use dropna when the number is small or insignificant
Netflix_df['director'].fillna('No director',inplace=True)
Netflix_df['cast'].fillna('No cast',inplace=True)
Netflix_df['country'].fillna(0,inplace=True)
#Netflix_df.dropna('date_added',axis=1)
#Netflix_df.dropna(subset=['date_added','rating','duration'], inplace=True)
```

```
Netflix_df.head(5)
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listec
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	No cast	United States	September 25, 2021	2020	PG-13	90 min	Documentar
1	s2	TV Show	Blood & Water	No director	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	Internatio TV Shows, Dramas, Myster
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	0	September 24, 2021	2021	TV-MA	1 Season	Crime Sho Internatio TV Shows, Ai
3	s4	TV Show	Jailbirds New Orleans	No director	No cast	0	September 24, 2021	2021	TV-MA	1 Season	Docuser Reality
4	s5	TV Show	Kota Factory	No director	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	Internatio TV Sho Romantic Shows, T

```
Netflix_df.isnull().values.any()
```

False

```
Netflix_df.isnull().any()
```

```
show_id      False
type         False
title        False
director     False
cast         False
country      False
```

```

date_added      False
release_year    False
rating          False
duration        False
listed_in       False
description     False
dtype: bool

```

```
print(Netflix_df.size, Netflix_df.shape)
```

105480 (8790, 12)

## Splitting the dataset into two according to type(movies/ Tv shows)

```
Netflix_df.head(2)
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_i
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	No cast	United States	September 25, 2021	2020	PG-13	90 min	Documentarie
1	s2	TV Show	Blood & Water	No director	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, T Dramas, T Mysterie

```

# Splitting the dataset according to type (Movies and Tv show)
# For Movies
Netflix_Movie =Netflix_df[Netflix_df['type']=='Movie']
Netflix_Movie.head(4)

```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	list
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	No cast	United States	September 25, 2021	2020	PG-13	90 min	Documen
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	0	September 24, 2021	2021	PG	91 min	Child Family M

show_id	type		title	director	cast	country	date_added	release_year	rating	duration	list
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	Dr: Indepe: M Internat: M
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG-13	104 min	Com Dr:

*# For TV shows*

```
Netflix_TV= Netflix_df[Netflix_df['type']=='TV Show']
Netflix_TV.head(4)
```

show_id	type		title	director	cast	country	date_added	release_year	rating	duration	listed_in
1	s2	TV Show	Blood & Water	No director	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	0	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
3	s4	TV Show	Jailbirds New Orleans	No director	No cast	0	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV
4	s5	TV Show	Kota Factory	No director	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...

*# Separating season 1 and 2 from the TV show*

```
Netflix_Season1= Netflix_df[Netflix_df['duration']=='1 Season']
```

```
Netflix_Season1.head(2)
```

show_id	type		title	director	cast	country	date_added	release_year	rating	duration	listed_in
---------	------	--	-------	----------	------	---------	------------	--------------	--------	----------	-----------

show_id	type		title	director	cast	country	date_added	release_year	rating	duration	listed_in
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	0	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
3	s4	TV Show	Jailbirds New Orleans	No director	No cast	0	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV

```
Netflix_Season2= Netflix_df[Netflix_df['duration']=='2 Seasons'] # separating season 2
Netflix_Season2.head(2)
```

show_id	type		title	director	cast	country	date_added	release_year	rating	duration	listed_in	de
1	s2	TV Show	Blood & Water	No director	Ama Qamata, Khosi Ngema, Gail Mabalan...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	p
4	s5	TV Show	Kota Factory	No director	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	Ir

## Rename a specific column or columns

```
Netflix_TV.rename(columns={'duration':'season'}, inplace=True) # To rename a column head
#Netflix_Movie.duration= Netflix_Movie.duration.str.replace('min','') # To remove string
Netflix_TV.rename(columns={'mad':'country'},inplace=True)
Netflix_TV.rename(columns={'season':'Seasons'}, inplace= True) # or
#Netflix_TV.replace({'Season':{'1 Season':'2 Seasons'}}, inplace= True)
```

/home/miracle/anaconda3/lib/python3.8/site-packages/pandas/core/frame.py:4441:

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)

```
return super().rename(
```

```
Netflix_TV.head(4)
```

show_id	type		title	director	cast	country	date_added	release_year	rating	Seasons	listed_in
---------	------	--	-------	----------	------	---------	------------	--------------	--------	---------	-----------



show_id	type		title	director	cast	country	date_added	release_year	rating	Seasons	listed_in
1	s2	TV Show	Blood & Water	No director	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	0	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
3	s4	TV Show	Jailbirds New Orleans	No director	No cast	0	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV
4	s5	TV Show	Kota Factory	No director	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...

## Replace specific characters in a columns

```
Netflix_TV.Seasons= Netflix_TV.Seasons.str.replace('2 Seasons', '2 Season')#Replace Spec
```

/home/miracle/anaconda3/lib/python3.8/site-packages/pandas/core/generic.py:5494:

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)

```
self[name] = value
```

```
Netflix_TV.Seasons=Netflix_TV.Seasons.str.replace('Season', '')
```

```
Netflix_TV.head(4)
```

show_id	type		title	director	cast	country	date_added	release_year	rating	Seasons	listed_in
1	s2	TV Show	Blood & Water	No director	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2	International TV Shows, TV Dramas, TV Mysteries

show_id	type		title	director	cast	country	date_added	release_year	rating	Seasons	listed_in
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	0	September 24, 2021	2021	TV-MA	1	Crime TV Shows, International TV Shows, TV Act...
3	s4	TV Show	Jailbirds New Orleans	No director	No cast	0	September 24, 2021	2021	TV-MA	1	Docuseries, Reality TV
4	s5	TV Show	Kota Factory	No director	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2	International TV Shows, Romantic TV Shows, TV ...

## Data Visualization

```
Netflix_df.head(5)
```

show_id	type		title	director	cast	country	date_added	release_year	rating	duration	listec
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	No cast	United States	September 25, 2021	2020	PG-13	90 min	Documenta
1	s2	TV Show	Blood & Water	No director	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	Internatio TV Shows, Dramas, Myster
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	0	September 24, 2021	2021	TV-MA	1 Season	Crime Sho Internatio TV Shows, Ai
3	s4	TV Show	Jailbirds New Orleans	No director	No cast	0	September 24, 2021	2021	TV-MA	1 Season	Docuser Reality
4	s5	TV Show	Kota Factory	No director	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	Internatio TV Sho Romantic Shows, T

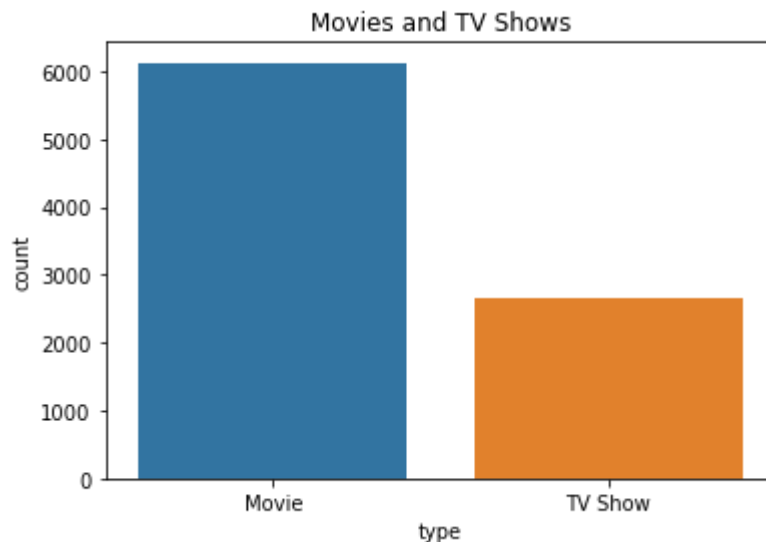
```
#sns.countplot(x='type', data=Netflix_df)
sns.countplot(Netflix_df.type)
plt.title('Movies and TV Shows')
```

/home/miracle/anaconda3/lib/python3.8/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.

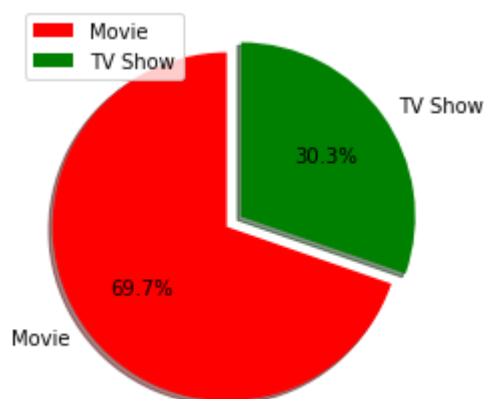
warnings.warn(

Text(0.5, 1.0, 'Movies and TV Shows')



```
plt.pie(Netflix_df.type.value_counts(),
        labels=Netflix_df.type.value_counts().index,
        explode=(0.05,0.05), colors=('r', 'g'),
        autopct='%1.1f%%', startangle=90, shadow=True);
plt.legend()
```

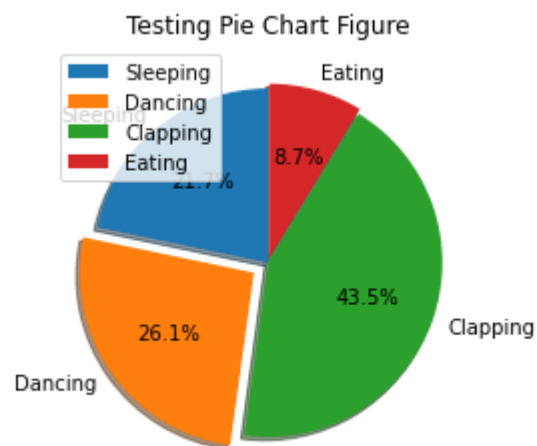
<matplotlib.legend.Legend at 0x7f3bf79af610>



```
# Testing the Pie chart.
Slices=[5,6,10,2]
activities=['Sleeping', 'Dancing', 'Clapping', 'Eating']
plt.pie(Slices, labels=activities, startangle=90,
        explode=(0,0.1,0,0.028), shadow=True, autopct='%0.1f%%')
```

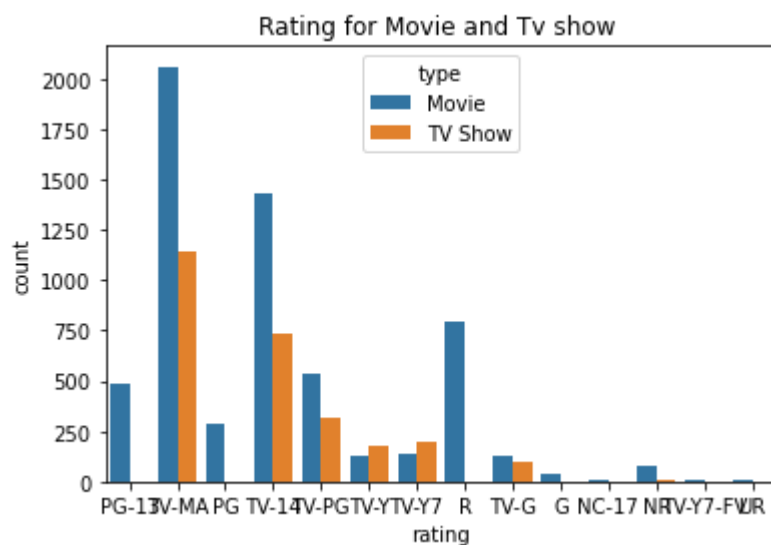
```
plt.legend()
plt.title('Testing Pie Chart Figure')
```

```
Text(0.5, 1.0, 'Testing Pie Chart Figure')
```

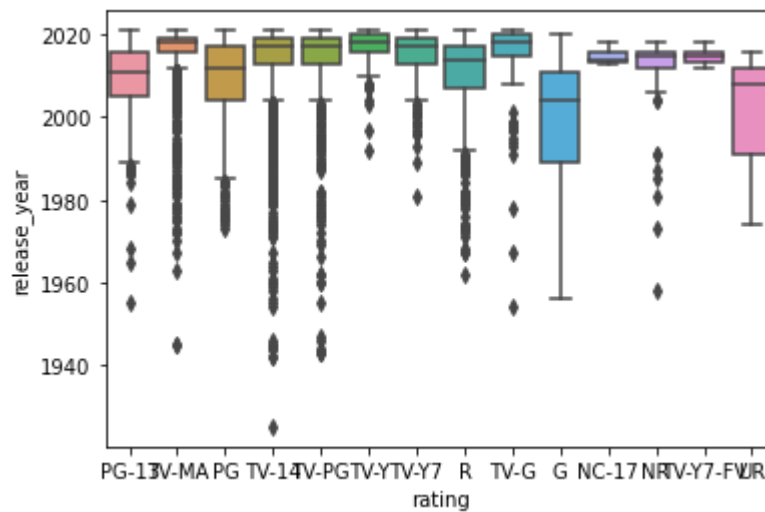


```
oder= ['G', 'TV-Y', 'TV-G', 'PG', 'TV-Y7', 'TV-Y7-FV', 'TV-PG', 'PG-13', 'TV-14', 'R', 'NC-17',
sns.countplot(x='rating', data=Netflix_df, hue='type')
#plt.figure(figsize=(15,7))
plt.title('Rating for Movie and Tv show ')
```

```
Text(0.5, 1.0, 'Rating for Movie and Tv show ')
```



```
sns.boxplot(x=Netflix_df['rating'], y=Netflix_df['release_year'])
plt.figure(figsize=(5,1))
plt.show()
```



<Figure size 360x72 with 0 Axes>

```
# Subplot Plotting
fig,ax=plt.subplots(1,2,figsize=(15,5))
g1=sns.countplot(Netflix_Movie.rating,ax=ax[0])
g1.set_xlabel('Rating')
g1.set_title('Rating for Movies')
g2=sns.countplot(Netflix_TV.rating,ax=ax[1])
g2.set_xlabel('Rating')
g2.set_title('Rating for TV Shows')
#g2.set(yticks=np.arange(0,1600,200))
```

/home/miracle/anaconda3/lib/python3.8/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.

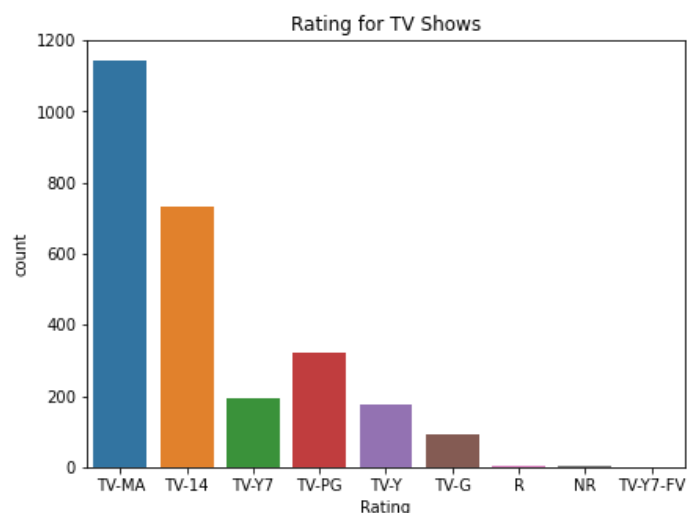
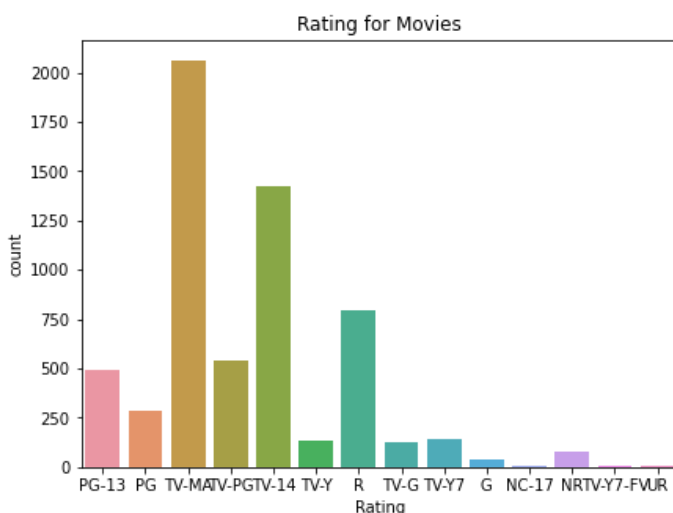
warnings.warn(

/home/miracle/anaconda3/lib/python3.8/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(

Text(0.5, 1.0, 'Rating for TV Shows')



# Counting the no of content added at release\_year columns

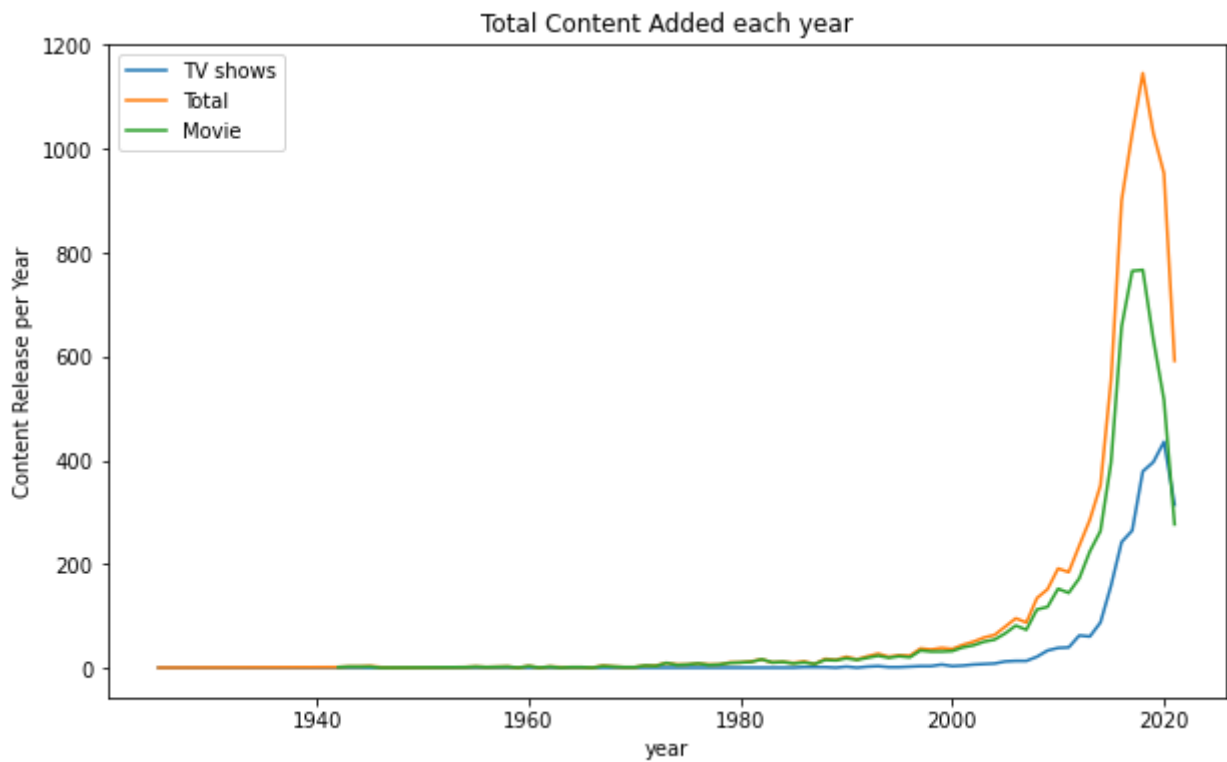
```
TV= Netflix_TV['release_year'].value_counts()
TV=pd. DataFrame(TV)
TV_count=TV.reset_index().rename(columns={'index':"year", 'release_year':'content_added'})
TV_count.head(5)
```

	year	content_added
0	2020	436
1	2019	397
2	2018	379
3	2021	315
4	2017	265

```
Movie= Netflix_Movie['release_year'].value_counts()
Movie_count= pd.DataFrame(Movie)
#Movie_count.columns=['year']
Movie_count= Movie_count.reset_index().rename(columns={'index':'year', 'release_year':'content_added'})
NetflixT= Netflix_df['release_year'].value_counts()
NetflixT_count=pd.DataFrame(NetflixT)
NetflixT_count=NetflixT_count.reset_index().rename(columns={'index':'year', 'release_year':'content_added'})
#NetflixT_count
```

```
plt.figure(figsize=(10,6))
sns.lineplot(x='year',y ='content_added',data=TV_count,label='TV shows')
sns.lineplot(x='year',y ='content_added',data=NetflixT_count, label='Total')
sns.lineplot(x='year',y ='content_added',data=Movie_count, label='Movie')
plt.ylabel('Content Release per Year')
plt.title('Total Content Added each year')
```

```
Text(0.5, 1.0, 'Total Content Added each year')
```



## How to sort date in months, days and years

```
Netflix_df['days']=pd.DatetimeIndex(Netflix_df['date_added']).day
Netflix_df['month_added']=pd.DatetimeIndex(Netflix_df['date_added']).month
Netflix_df['year']=pd.DatetimeIndex(Netflix_df['date_added']).year
Netflix_df['miracle']=TV
Netflix_df.head(3)
```

**NameError**

Traceback (most recent call last)

<ipython-input-1-f1ddb3ef7fac> in <module>

```
----> 1 Netflix_df['days']=pd.DatetimeIndex(Netflix_df['date_added']).day
      2 Netflix_df['month_added']=pd.DatetimeIndex(Netflix_df['date_added']).month
      3 Netflix_df['year']=pd.DatetimeIndex(Netflix_df['date_added']).year
      4 Netflix_df['miracle']=TV
      5 Netflix_df.head(3)
```

**NameError**: name 'pd' is not defined

```
project_name='Netflix_data_analysis'
jovian .commit(project=project_name)
```