

# Exploratory Data Analysis of Netflix Movies and Tv Shows.

Netflix, in full Netflix, Inc., media-streaming and video-rental company founded in 1997 by American entrepreneurs Reed Hastings and Marc Randolph. It is also involved in the creation of original programming. Corporate headquarters are in Los Gatos, California. In 1999 Netflix began offering an online subscription service through the Internet. Subscribers chose movie and television titles from Netflix's Web site; the shows were then mailed to customers in the form of DVDs, along with prepaid return envelopes, from one of more than 100 distribution centres. Although customers typically rented for a flat monthly fee as many movies per month as they wished, the number of DVDs in their possession at any one time was limited according to their subscription plans. Netflix had tens of thousands of movie titles in its catalog. and is you find more information about Netflix ne click this <https://www.britannica.com/topic/Netflix-Inc>. Netflix is one of the most popular media and video streaming platforms. They have over 9500 movies or tv shows available on their platform, as of mid-2021, they have over 250M Subscribers in all over countries. This tabular dataset consists of listings of all the movies and tv shows available on Netflix, along with details such as - cast, directors, ratings, release year, duration, etc

And in this Netflix datasets, all details include like-cast, release year, actors, director, yearly data, movie title etc, And this dataset found from kaggle website <https://www.kaggle.com/datasets/shivamb/netflix-shows>, And we will do details Analysis of netflix movies and shows like highest release movie year, trending movie, and cleaning, visualization, and also give question- answers and we use in this Analysis tools like- Pandas, Matplotlib, seaborn, numpy etc. and all activity done in course [Data Analysis with Python: Zero to Pandas](#),

And we follow following step like-

- Dataset download from online source
- Data Preparation and Cleaning
- Exploratory Analysis and Visualization
- Asking and Answering Questions
- Inferences and Conclusion
- References and Future Work

## Downloading the Dataset

First step is Download the Dataset, in this step we will download Netflix dataset from online source of Kaggle website <https://www.kaggle.com/datasets/shivamb/netflix-shows>

```
!pip install jovian opendatasets --upgrade --quiet
```

Let's begin by downloading the data, and listing the files within the dataset.

```
# Change this
dataset_url = 'https://www.kaggle.com/datasets/shivamb/netflix-shows'
```

```
import opendatasets as od
od.download(dataset_url)
```

Please provide your Kaggle credentials to download this dataset. Learn more:

<http://bit.ly/kaggle-creds>

Your Kaggle username: anilmodi123

Your Kaggle Key: .....

Downloading netflix-shows.zip to ./netflix-shows

100%|██████████| 1.34M/1.34M [00:00<00:00, 65.1MB/s]

The dataset has been downloaded and extracted.

```
data_dir = './netflix-shows'
```

```
import os
os.listdir(data_dir)
```

```
['netflix_titles.csv']
```

```
project_name = "netflix movies and tv shows"
```

```
!pip install jovian --upgrade -q
```

```
import jovian
```

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

[jovian] Updating notebook "amodi1486/netflix-movies-and-tv-shows" on <https://jovian.ai>

[jovian] Committed successfully! <https://jovian.ai/amodi1486/netflix-movies-and-tv-shows>

'<https://jovian.ai/amodi1486/netflix-movies-and-tv-shows>'

## Data Preparation and Cleaning

Data cleaning is the process of editing, correcting, and structuring data within a data set so that it's generally uniform and prepared for analysis. This includes removing or replace missing value or irrelevant data and formatting it into a language that computers can understand for optimal analysis

Now we comming on Second step Data Preparation and Cleaning. in this step we we done clear datasets, like

- First of all, we will done import pandas for read csv file
- Finding Information about Dataset
- Finding Columns, shape and dtype of Datasets
- Change data type
- Finding duplicate and null values in datasetsfind

- And filling missing value and drop some rows in Dataset

```
import pandas as pd
```

```
Netflix_df = pd.read_csv('netflix-shows/netflix_titles.csv')
```

Netflix\_df

| show_id |       | type    | title                 | director        | cast   | country       | date_added         | release_year | rating | duration  |                   |
|---------|-------|---------|-----------------------|-----------------|--|---------------|--------------------|--------------|--------|-----------|-------------------|
| 0       | s1    | Movie   | Dick Johnson Is Dead  | Kirsten Johnson | NaN  | United States | September 25, 2021 | 2020         | PG-13  | 90 min    | Docum             |
| 1       | s2    | TV Show | Blood & Water         | NaN             | Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...  | South Africa  | September 24, 2021 | 2021         | TV-MA  | 2 Seasons | Inter TV Si Drz I |
| 2       | s3    | TV Show | Ganglands             | Julien Leclercq | Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabil... | NaN           | September 24, 2021 | 2021         | TV-MA  | 1 Season  | ( Inter TV Si     |
| 3       | s4    | TV Show | Jailbirds New Orleans | NaN             | NaN  | NaN           | September 24, 2021 | 2021         | TV-MA  | 1 Season  | Do R              |
| 4       | s5    | TV Show | Kota Factory          | NaN             | Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...  | India         | September 24, 2021 | 2021         | TV-MA  | 2 Seasons | Inter T' Rom Sho  |
| ...     | ...   | ...     | ...                   | ...             | ...  | ...           | ...                | ...          | ...    | ...       |                   |
| 8802    | s8803 | Movie   | Zodiac                | David Fincher   | Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...  | United States | November 20, 2019  | 2007         | R      | 158 min   | Cul               |
| 8803    | s8804 | TV Show | Zombie Dumb           | NaN             | NaN  | NaN           | July 1, 2019       | 2018         | TV-Y7  | 2 Seasons | Ki Sl C           |

|      | show_id | type  | title      | director        | cast  | country       | date_added       | release_year | rating | duration |                   |
|------|---------|-------|------------|-----------------|---|---------------|------------------|--------------|--------|----------|-------------------|
| 8804 | s8805   | Movie | Zombieland | Ruben Fleischer | Jesse Eisenberg, Woody Harrelson, Emma Stone, ... | United States | November 1, 2019 | 2009         | R      | 88 min   | C Horrc           |
| 8805 | s8806   | Movie | Zoom       | Peter Hewitt    | Tim Allen, Courteney Cox, Chevy Chase, Kate Ma... | United States | January 11, 2020 | 2006         | PG     | 88 min   | C Famil<br>C      |
| 8806 | s8807   | Movie | Zubaan     | Moze Singh      | Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan... | India         | March 2, 2019    | 2015         | TV-14  | 111 min  | Inter<br>Mov<br>& |

8807 rows × 12 columns

The dataset contains over 8807 rows and 12 columns. this dataset collected from 1943 to 2021. And this dataset contains all informatio like- type of movies, cast, title, directors,counytry ,date added on Netflix, release date, genres etc..

# Information

```
Netflix_df.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 this store all information like how many data store in perticular columns, in int and object datatype

# Columns ,shape and dtype

```
Netflix_df.columns
```

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

All columns include in this Datasets

## Change Dtype of Date column

```
Netflix_df['date_added'] = pd.to_datetime(Netflix_df['date_added'])
```

```
Netflix_df.dtypes
```

```
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  
dtype: object
```

```
Netflix_df.shape
```

```
(8807, 12)
```

In this datasets include 8807 rows and 12 column.

```
Netflix_df.dtypes
```

```
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
dtype: object
```

IN this datasets all datatype is object but release\_year datatype is only int type

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

|   | show_id | type    | title                 | director        | cast  | country       | date_added | release_year | rating | duration  | listed_in                           |
|---|---------|---------|-----------------------|-----------------|---|---------------|------------|--------------|--------|-----------|-------------------------------------|
| 0 | s1      | Movie   | Dick Johnson Is Dead  | Kirsten Johnson | NaN   | United States | 2021-09-25 | 2020         | PG-13  | 90 min    | Documentar                          |
| 1 | s2      | TV Show | Blood & Water         | NaN             | Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban... | South Africa  | 2021-09-24 | 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           | 2021-09-24 | 2021         | TV-MA  | 1 Season  | Crime Sho Internatio TV Shows, Ai   |
| 3 | s4      | TV Show | Jailbirds New Orleans | NaN             | NaN   | NaN           | 2021-09-24 | 2021         | TV-MA  | 1 Season  | Docuser Reality                     |
| 4 | s5      | TV Show | Kota Factory          | NaN             | Mayur More, Jitendra Kumar, Ranjan Raj, Alam K... | India         | 2021-09-24 | 2021         | TV-MA  | 2 Seasons | Internatio TV Sho Romantic Shows, T |

## Check Duplicate and null values

```
Netflix_df[Netflix_df.duplicated()]
```

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

In this Data sets no Any Duplicate value

```
Netflix_df.isnull().sum().sort_values(ascending=False)
```

```
director      2634
country        831
cast           825
date_added     10
```

```
rating          4
duration        3
show_id         0
type            0
title           0
release_year    0
listed_in       0
description      0
dtype: int64
```

- Highest missing value in Director, next country and cast column
- Least replace/ filling the missing value

```
Netflix_df.director.fillna('No Director', inplace=True)
Netflix_df.cast.fillna('No Cast', inplace=True)
Netflix_df.country.fillna('Country Unavailable', inplace=True)

Netflix_df.dropna(subset=['date_added', 'rating'], inplace=True)
```

In this dataset missing value in director, cast and country,

- missing value in Director column filled done with ('No Director')
- missing value in Cast column filled done with ('No Cast')
- missing value in Country column filled with ('No Country')
- and next 2 column date\_added, and rating, we must need to fix value for fill missing value but we have no any fix value, so drop the missing values in the Dataset of Netflix

```
Netflix_df.isnull().sum()
```

```
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
```

```
import jovian
```

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

```
[jovian] Updating notebook "amodi1486/netflix-movies-and-tv-shows" on https://jovian.ai  
[jovian] Committed successfully! https://jovian.ai/amodi1486/netflix-movies-and-tv-shows  
'https://jovian.ai/amodi1486/netflix-movies-and-tv-shows'
```

## Exploratory Analysis and Visualization

Exploratory Data Analysis refers to the critical process of performing initial Analysis on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations. we will Analysis following columns from Dataset.

- First of all, we import seaborn, matplotlib, matplotlib.pyplot and set a fix parameter
- Type
- Release\_year
- listed\_in (Genres)
- Rating
- cast (Actors)
- Country

Let's begin by importing matplotlib.pyplot and seaborn .

```
import seaborn as sns  
import matplotlib  
import matplotlib.pyplot as plt  
%matplotlib inline  
  
sns.set_style('darkgrid')  
matplotlib.rcParams['font.size'] = 14  
matplotlib.rcParams['figure.figsize'] = (9, 5)  
matplotlib.rcParams['figure.facecolor'] = '#00000000'
```

## Type

```
Netflix_df.groupby(['type']).type.count()
```

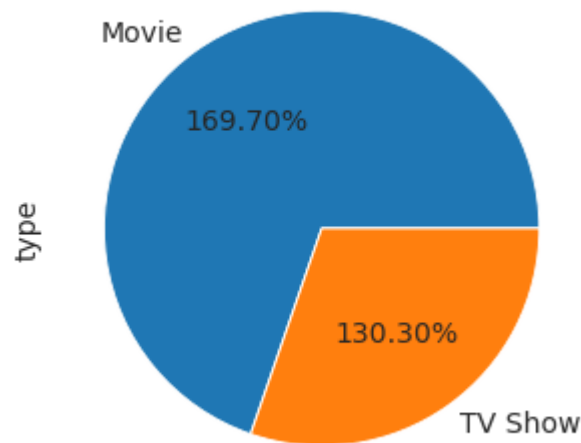
```
type  
Movie      6129  
TV Show    2664  
Name: type, dtype: int64
```

In this Datasets, total movie is 6131 and tv show is 2676,

```
Netflix_df['type'].value_counts().plot(kind='pie', autopct='1%.2f%%')  
plt.title('Netflix Movies and Tv shows');
```



## Netflix Movies and Tv shows



In this type columns, we found number of movies is highest release than compare to Tv shows

- number of movies= 169.70 %
- number of tv shows = 130.30%

## Release\_year

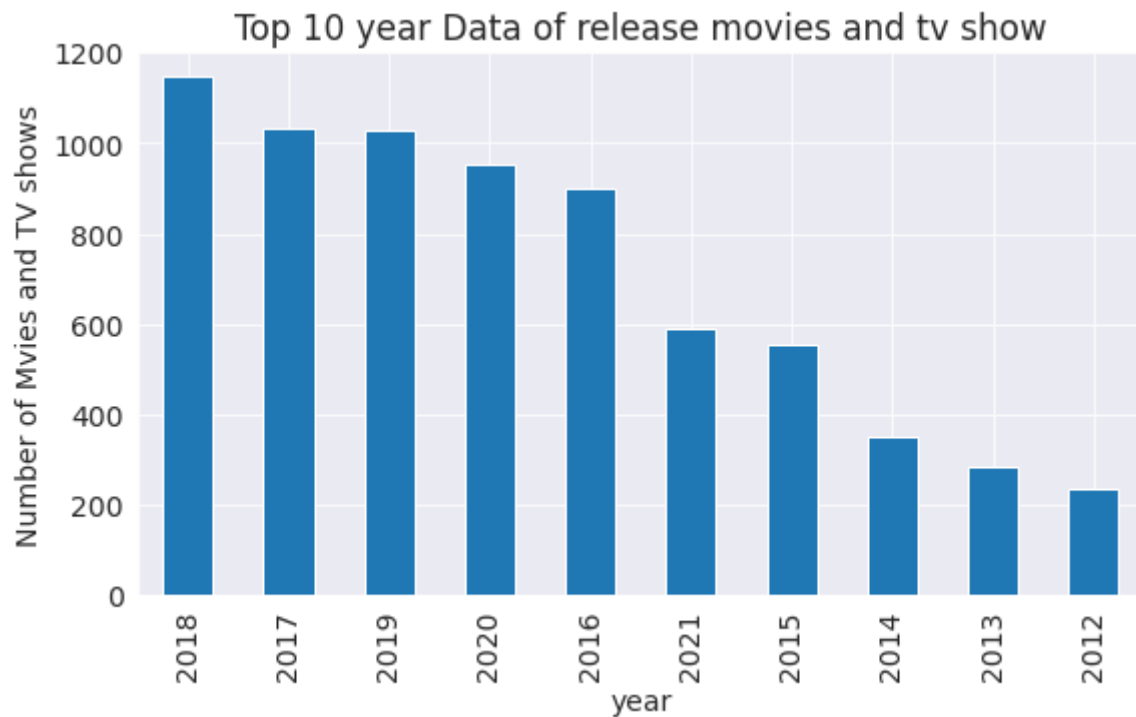
```
netflix_release_year = Netflix_df.release_year.value_counts().sort_values(ascending=False)
```

```
netflix_release_year
```

```
2018    1146
2017    1031
2019    1030
2020     953
2016     901
...
1959      1
1925      1
1961      1
1947      1
1966      1
```

```
Name: release_year, Length: 74, dtype: int64
```

```
netflix_release_year[:10].plot(kind='bar')
plt.xlabel('year')
plt.ylabel('Number of Mvies and TV shows')
plt.title('Top 10 year Data of release movies and tv show');
```



In this chart we find that the highest movies and TV show release in 2018 on Netflix.

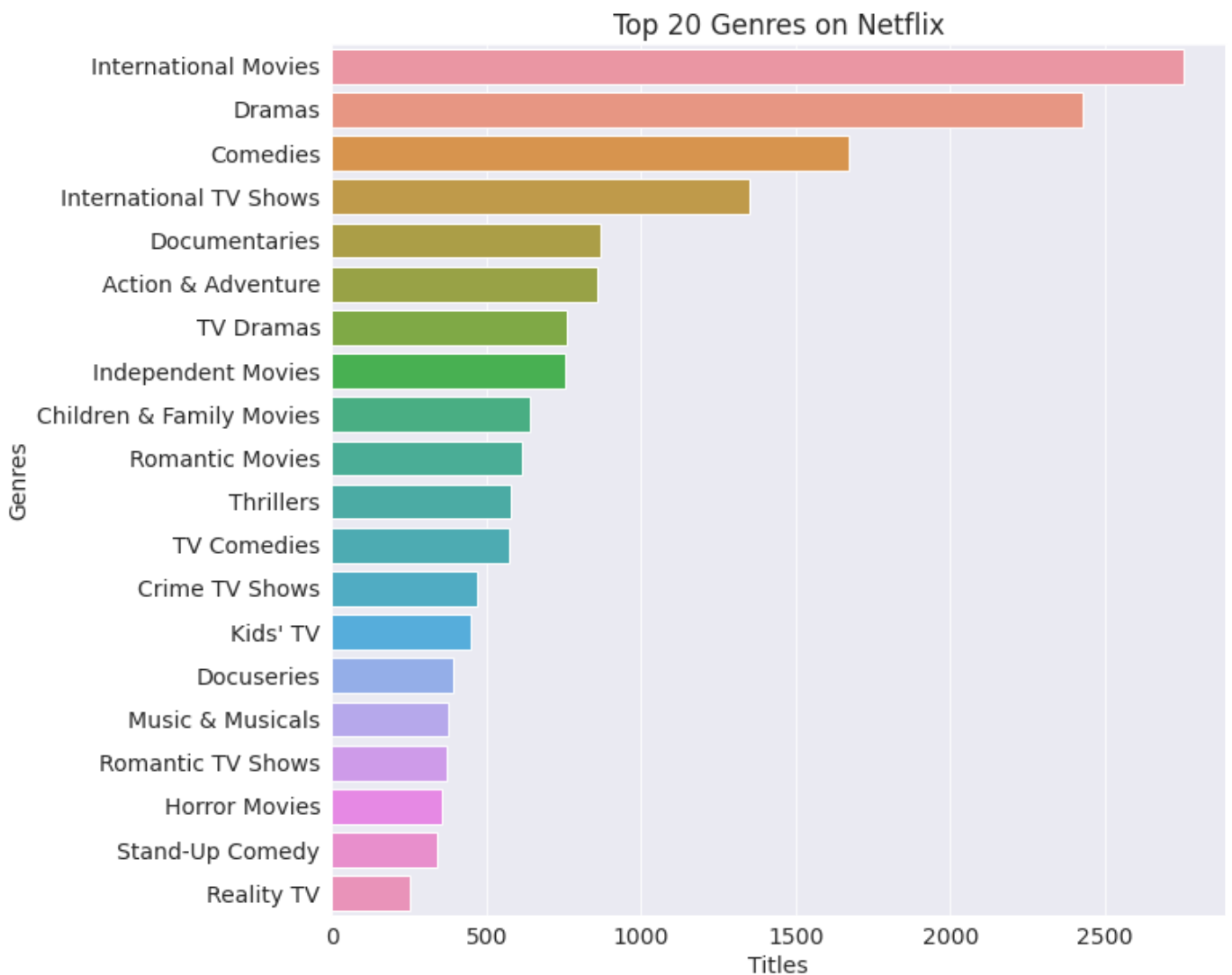
- total movies and tv shows release in 2018 = 1146

## listed\_in ( Genres)

```
filtered_genres = Netflix_df.set_index('title').listed_in.str.split(', ', expand=True).
filtered_genres
```

```
title
Dick Johnson Is Dead      Documentaries
Blood & Water             International TV Shows
Blood & Water             TV Dramas
Blood & Water             TV Mysteries
Ganglands                Crime TV Shows
...
Zoom                     Children & Family Movies
Zoom                     Comedies
Zubaan                   Dramas
Zubaan                   International Movies
Zubaan                   Music & Musicals
Length: 19297, dtype: object
```

```
plt.figure(figsize=(10,10))
g = sns.countplot(y = filtered_genres, order=filtered_genres.value_counts().index[:20])
plt.title('Top 20 Genres on Netflix')
plt.xlabel('Titles')
plt.ylabel('Genres')
plt.show()
```



- In Genres, International movies are most release on Netflix,
- Next Darama also big number of release on Netflix,
- and minimum number of Reality TV shows are release on Netflix

## Rating

```
Netflix_df.rating.unique()
```

```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
      'TV-G', 'G', 'NC-17', '74 min', '84 min', '66 min', 'NR',
      'TV-Y7-FV', 'UR'], dtype=object)
```

```
Netflix_rating = Netflix_df.set_index('title').rating.str.split(', ', expand=True).stack()
Netflix_rating
```

| title                 |       |
|-----------------------|-------|
| Dick Johnson Is Dead  | PG-13 |
| Blood & Water         | TV-MA |
| Ganglands             | TV-MA |
| Jailbirds New Orleans | TV-MA |
| Kota Factory          | TV-MA |

```

...
Zodiac R
Zombie Dumb TV-Y7
Zombieland R
Zoom PG
Zubaan TV-14
Length: 8793, dtype: object

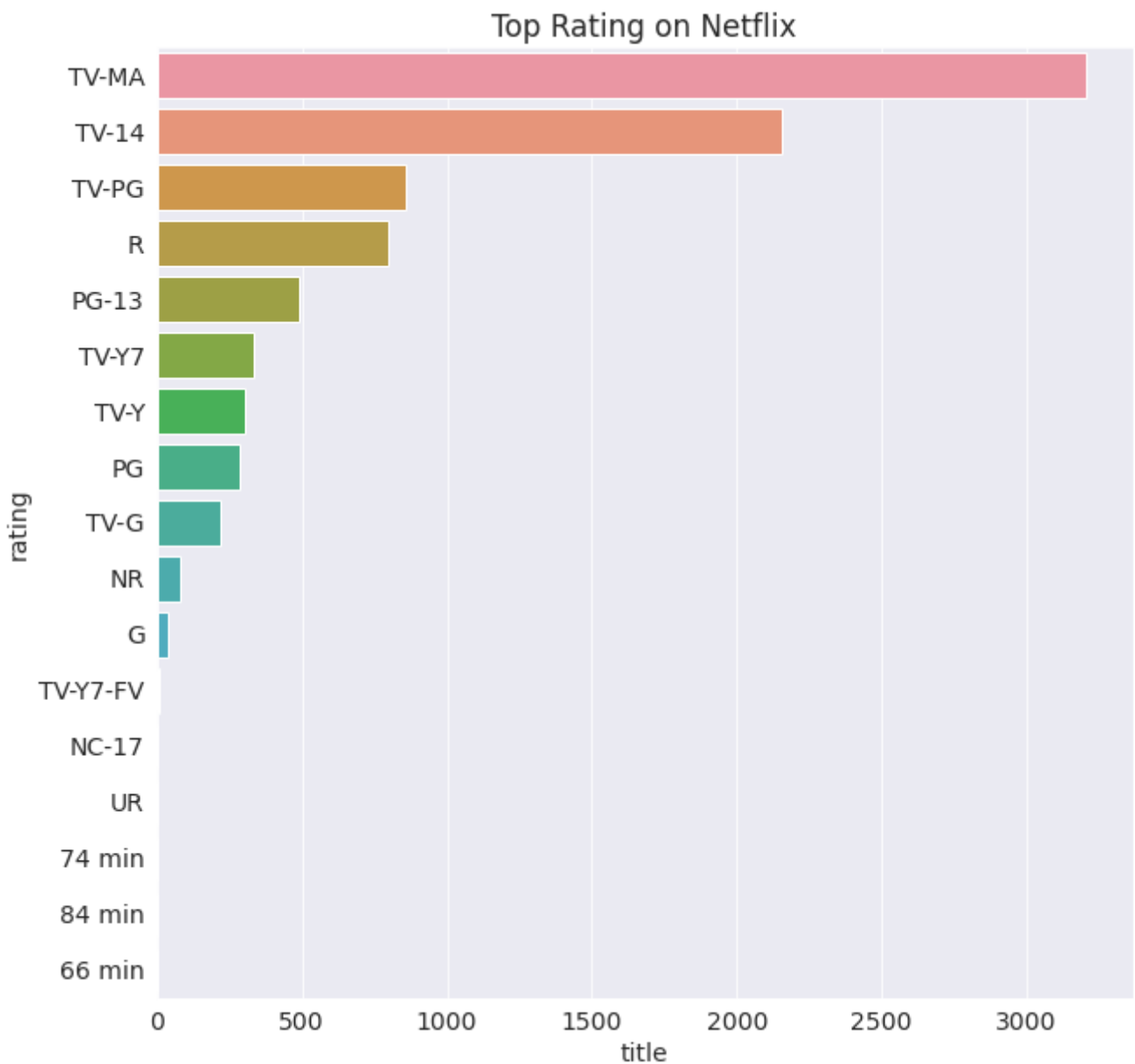
```

```

plt.figure(figsize=(10,10))
sns.countplot(y=Netflix_rating, order=Netflix_rating.value_counts().index[:20])
plt.xlabel('title')
plt.ylabel('rating')
plt.title('Top Rating on Netflix')

plt.show()

```



In the Rating in Dataset, Highest Rating is TV-MA on Netflix

- TV-MA For **Mature Audiences**. May not be suitable for ages 17 and under
- and next is TA-14 rating high on Netflix

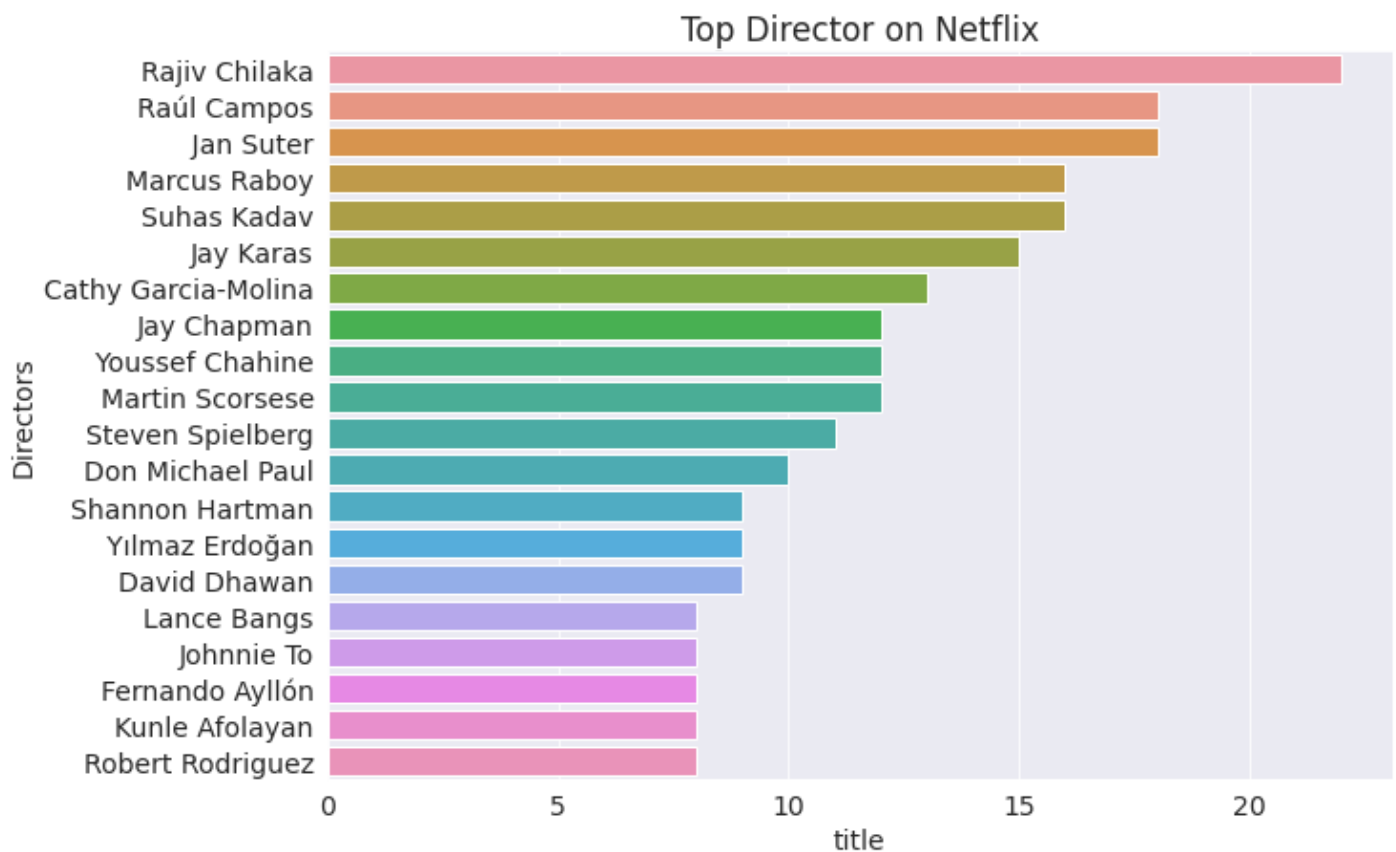
- Tv-14 Parents strongly cautioned. May not be suitable for ages 14 and under.

## Director

```
Netflix_director = Netflix_df.set_index('title').director.str.split(',', expand=True).stack()
Netflix_director
```

```
title
Dick Johnson Is Dead      Kirsten Johnson
Blood & Water             No Director
Ganglands                 Julien Leclercq
Jailbirds New Orleans     No Director
Kota Factory              No Director
...
Zodiac                    David Fincher
Zombie Dumb               No Director
Zombieland                Ruben Fleischer
Zoom                      Peter Hewitt
Zubaan                    Mozez Singh
Length: 9598, dtype: object
```

```
Netflix_director = Netflix_director[Netflix_director != 'No Director']
plt.figure(figsize=(10,7))
d = sns.countplot(y = Netflix_director, order=Netflix_director.value_counts().index[:20])
plt.xlabel('title')
plt.ylabel('Directors')
plt.title('Top Director on Netflix')
plt.show()
```



This Graph is shows Contributor of Director on Netflix

- In this Dataset, **Rajiv Chilaka** is highest movies and TV shows are release on Netflix
- and next **Raul Campos** and **Jan Suter** are also highes movies and Tv shows are release on Netflix

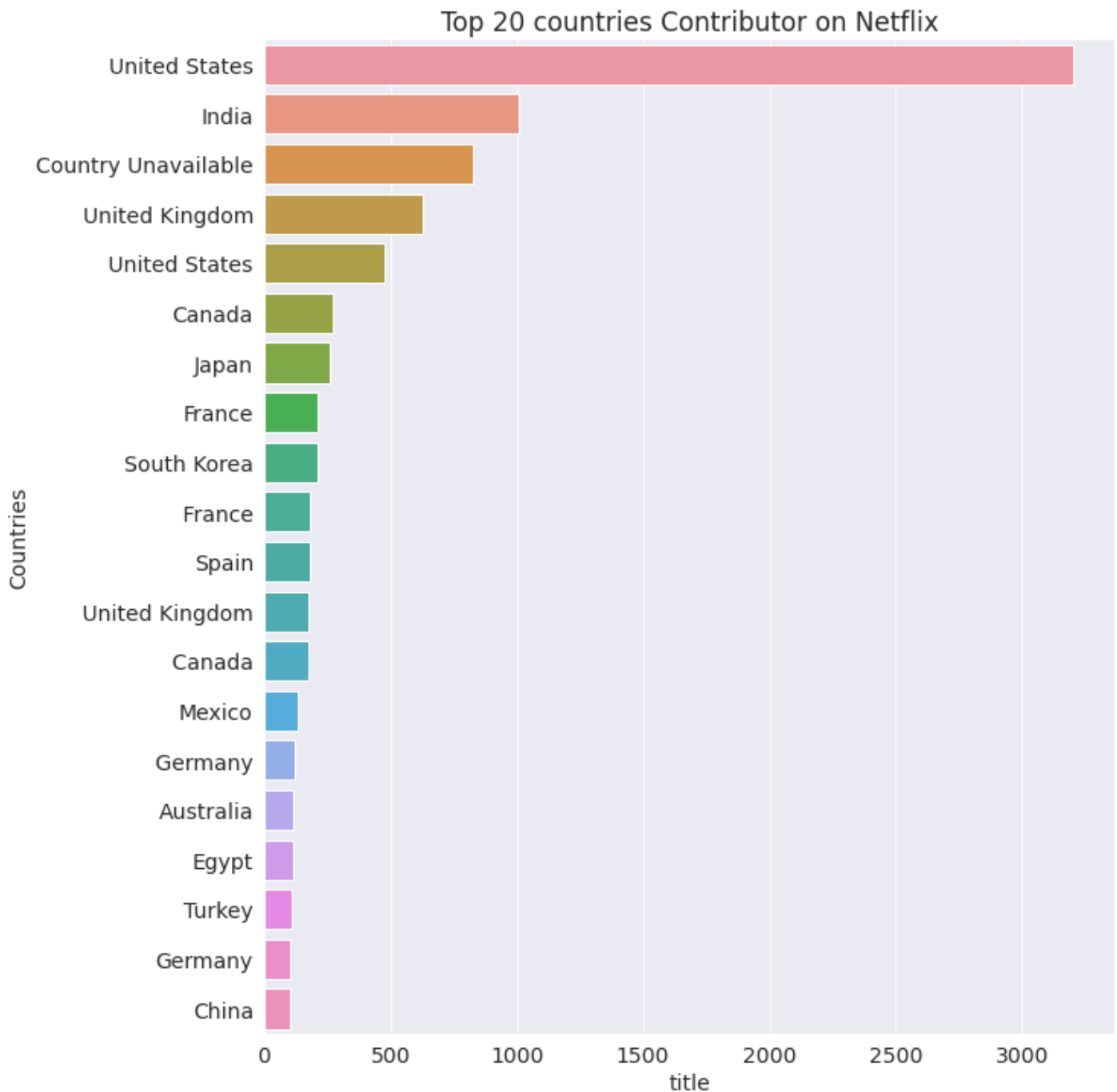
## Country

```
Netflix_countries = Netflix_df.set_index('title').country.str.split(',', expand=True).stack()
Netflix_countries
```

```
title
Dick Johnson Is Dead      United States
Blood & Water             South Africa
Ganglands                 Country Unavailable
Jailbirds New Orleans     Country Unavailable
Kota Factory              India
...
Zodiac                    United States
Zombie Dumb               Country Unavailable
Zombieland                United States
Zoom                      United States
Zubaan                    India
Length: 10836, dtype: object
```

```
plt.figure(figsize=(10,12))
sns.countplot(y=Netflix_countries, order=Netflix_countries.value_counts().index[:20])
plt.xlabel('title')
```

```
plt.ylabel('Countries')
plt.title('Top 20 countries Contributor on Netflix')
plt.show()
```



This Countplot shows Country Contributor on Netflix

- **United States** is most release of movies and tv shows on Netflix
- and next **India** is a second country contributor on Netflix

## Cast ( Actors)

```
Netflix_actors= Netflix_df.set_index('title').cast.str.split(', ', expand=True).stack().r
Netflix_actors
```

title

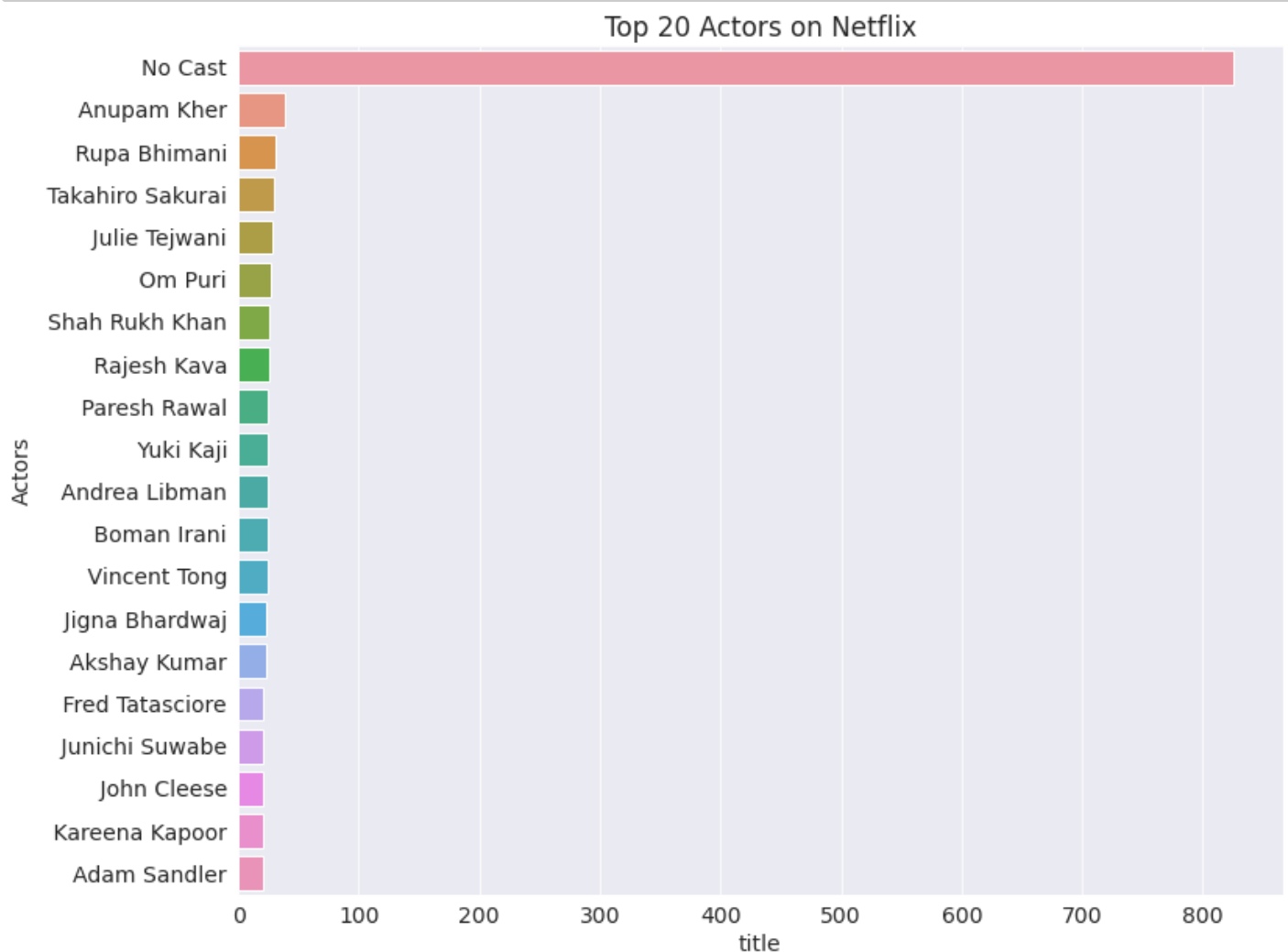
Dick Johnson Is Dead

No Cast

|               |                       |
|---------------|-----------------------|
| Blood & Water | Ama Qamata            |
| Blood & Water | Khosi Ngema           |
| Blood & Water | Gail Mabalane         |
| Blood & Water | Thabang Molaba        |
|               | ...                   |
| Zubaan        | Manish Chaudhary      |
| Zubaan        | Meghna Malik          |
| Zubaan        | Malkeet Rauni         |
| Zubaan        | Anita Shabdish        |
| Zubaan        | Chittaranjan Tripathy |

Length: 64844, dtype: object

```
plt.figure(figsize=(12,10))
sns.countplot(y=Netflix_actors, order=Netflix_actors.value_counts().index[:20])
Netflix_actors = Netflix_actors[Netflix_actors != 'No Cast']
plt.xlabel('title')
plt.ylabel('Actors')
plt.title('Top 20 Actors on Netflix')
plt.show()
```



This Graph shows Actors are trend on Netflix;

- The indian Actor **Anupam Kher** is the most popular on Netflix
- and next other like **Rupa Bhimani**, **Takahiro sakurai** and **om puri** are also popular on Netflix



```
import jovian
```

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

[jovian] Updating notebook "amodi1486/netflix-movies-and-tv-shows" on <https://jovian.ai>

[jovian] Committed successfully! <https://jovian.ai/amodi1486/netflix-movies-and-tv-shows>

'<https://jovian.ai/amodi1486/netflix-movies-and-tv-shows>'

## Asking and Answering Questions

In this step, we ask question about dataset of Netflix, and we found Answers by Analysis the Netflix Datasets, and visualization, following ask question in this dataset

- Which type of category has a larger number of releases in india ? explain by chart
- Which year release the maximum number of movies and tv shows in india ? explain by graph
- Which Genres play the most in india on Netflix ? explain by chart
- In how many movies / Tv shows, Anupam kher was cast ?
- Which month has the most content add on Netlix ? show with Graph !

**Q1: Which Type of category has a larger number of release in india ? explain by chart !**

```
Netflix_india = Netflix_df[Netflix_df['country']=='India']  
Netflix_india.head(5)
```

| show_id | type           | title         | director    | cast  | country | date_added | release_year | rating | duration  | li                     |
|---------|----------------|---------------|-------------|---|---------|------------|--------------|--------|-----------|------------------------|
| 4       | s5<br>TV Show  | Kota Factory  | No Director | Mayur More, Jitendra Kumar, Ranjan Raj, Alam K... | India   | 2021-09-24 | 2021         | TV-MA  | 2 Seasons | Intern: TV: Ro: TV:    |
| 24      | s25<br>Movie   | Jeans         | S. Shankar  | Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi... | India   | 2021-09-21 | 1998         | TV-14  | 166 min   | Cor Intern: Ro: I      |
| 39      | s40<br>TV Show | Chhota Bheem  | No Director | Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig... | India   | 2021-09-16 | 2021         | TV-Y7  | 3 Seasons | K                      |
| 50      | s51<br>TV Show | Dharmakshetra | No Director | Kashmira Irani, Chandan Anand, Dinesh Mehta, A... | India   | 2021-09-15 | 2014         | TV-PG  | 1 Season  | Intern: TV: TV D TV Sc |

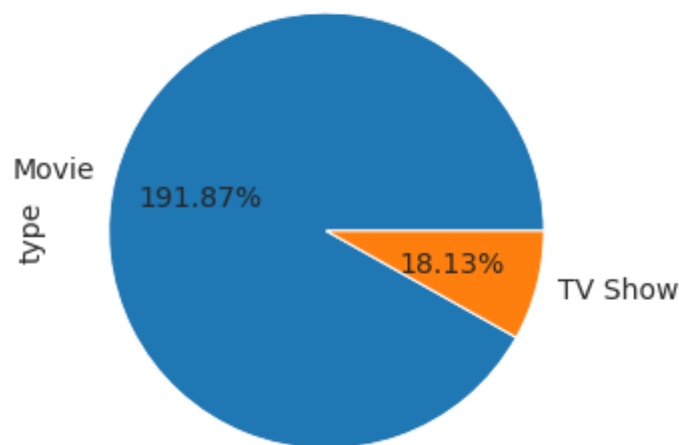
|    | show_id | type    | title                         | director    | cast    | country | date_added | release_year | rating | duration | li:             |
|----|---------|---------|-------------------------------|-------------|---------|---------|------------|--------------|--------|----------|-----------------|
| 66 | s67     | TV Show | Raja Rasoi Aur Anya Kahaniyan | No Director | No Cast | India   | 2021-09-15 | 2014         | TV-G   | 1 Season | Docu Intern: TV |

```
Netflix_india.type.value_counts()
```

```
Movie      893
TV Show    79
Name: type, dtype: int64
```

```
Netflix_india['type'].value_counts().plot(kind='pie', autopct='1%.2f%')
plt.title('Netflix Movies and Tv shows in india');
```

Netflix Movies and Tv shows in india



India is a biggest country, and Bollywood, Hindi-language sector of the Indian movie making industry that began in Bombay (now Mumbai) in the 1930s and developed into an enormous film empire. After early Indian experiments in silent film, in 1934 Bombay Talkies, launched by Himansu Rai, spearheaded the growth of Indian cinema. and now Boliwood come in top rank in release movies on Netflix. According to pie chart, now second highest number movies release of India on Netflix as compare to Tv show;

- Movies around = 191.87%
- and Tv shows around = 18.13%

**Q2: Which year release the maximum number of movies and tv shows in india ? explain by graph**

```
Netflix_india_df = Netflix_df[Netflix_df['country']=='India']
Netflix_india_df2= Netflix_india_df.groupby(['release_year'])['type'].value_counts().sort
Netflix_india_df2.head(10)
```

```
release_year  type
2017          Movie    98
2018          Movie    81
```

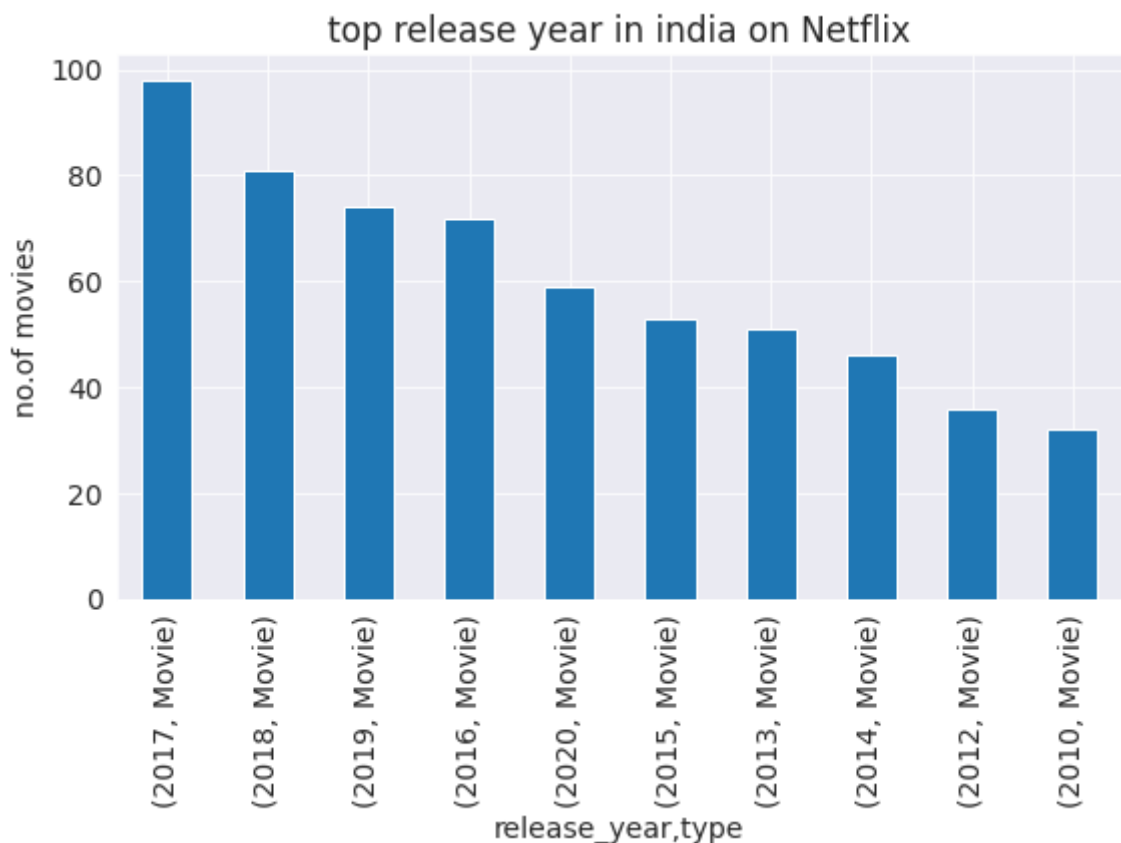
|      |       |    |
|------|-------|----|
| 2019 | Movie | 74 |
| 2016 | Movie | 72 |
| 2020 | Movie | 59 |
| 2015 | Movie | 53 |
| 2013 | Movie | 51 |
| 2014 | Movie | 46 |
| 2012 | Movie | 36 |
| 2010 | Movie | 32 |

Name: type, dtype: int64

```

Netlix_india_df2.head(10).plot(kind='bar');
plt.ylabel('no.of movies')
plt.title('top release year in india on Netflix');

```



According to Graph, highest only movies release on 2017 and 2018 on Netflix on bases of Netflix Dataset. after 2017, release movies are fall down because of covid 19,

Q3: Which type of Genres popular in india on Netflix ? show with Graph !

```

Netflix_india_ = Netflix_df[Netflix_df['country']=='India']
Netflix_india_

```

|  | show_id | type          | title        | director    | cast   | country | date_added | release_year | rating | duration  |    |
|--|---------|---------------|--------------|-------------|--|---------|------------|--------------|--------|-----------|----|
|  | 4       | s5<br>TV Show | Kota Factory | No Director | Mayur More,<br>Jitendra Kumar,<br>Ranjan Raj,<br>Alam K... | India   | 2021-09-24 | 2021         | TV-MA  | 2 Seasons | In |

|      | show_id | type    | title                         | director                | cast  | country | date_added | release_year | rating | duration  |        |
|------|---------|---------|-------------------------------|-------------------------|---|---------|------------|--------------|--------|-----------|--------|
| 24   | s25     | Movie   | Jeans                         | S. Shankar              | Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi... | India   | 2021-09-21 | 1998         | TV-14  | 166 min   | In     |
| 39   | s40     | TV Show | Chhota Bheem                  | No Director             | Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig... | India   | 2021-09-16 | 2021         | TV-Y7  | 3 Seasons |        |
| 50   | s51     | TV Show | Dharmakshetra                 | No Director             | Kashmira Irani, Chandan Anand, Dinesh Mehta, A... | India   | 2021-09-15 | 2014         | TV-PG  | 1 Season  | In - T |
| 66   | s67     | TV Show | Raja Rasoi Aur Anya Kahaniyan | No Director             | No Cast   | India   | 2021-09-15 | 2014         | TV-G   | 1 Season  | In     |
| ...  | ...     | ...     | ...                           | ...                     | ...   | ...     | ...        | ...          | ...    | ...       |        |
| 8773 | s8774   | Movie   | Yanda Kartavya Aahe           | Kedar Shinde            | Ankush Choudhary, Smita Shewale, Mohan Joshi, ... | India   | 2018-01-01 | 2006         | TV-PG  | 151 min   | In     |
| 8775 | s8776   | TV Show | Yeh Meri Family               | No Director             | Vishesh Bansal, Mona Singh, Akarsh Khurana, Ah... | India   | 2018-08-31 | 2018         | TV-PG  | 1 Season  | In     |
| 8798 | s8799   | Movie   | Zed Plus                      | Chandra Prakash Dwivedi | Adil Hussain, Mona Singh, K.K. Raina, Sanjay M... | India   | 2019-12-31 | 2014         | TV-MA  | 131 min   | In     |
| 8799 | s8800   | Movie   | Zenda                         | Avadhoot Gupte          | Santosh Juvekar, Siddharth Chandekar, Sachit P... | India   | 2018-02-15 | 2009         | TV-14  | 120 min   | In     |
| 8806 | s8807   | Movie   | Zubaan                        | Mozez Singh             | Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan... | India   | 2019-03-02 | 2015         | TV-14  | 111 min   | In     |

972 rows × 12 columns

```
Netflix_india_geners = Netflix_india_.set_index('title').listed_in.str.split(',', expand=True)
Netflix_india_geners
```

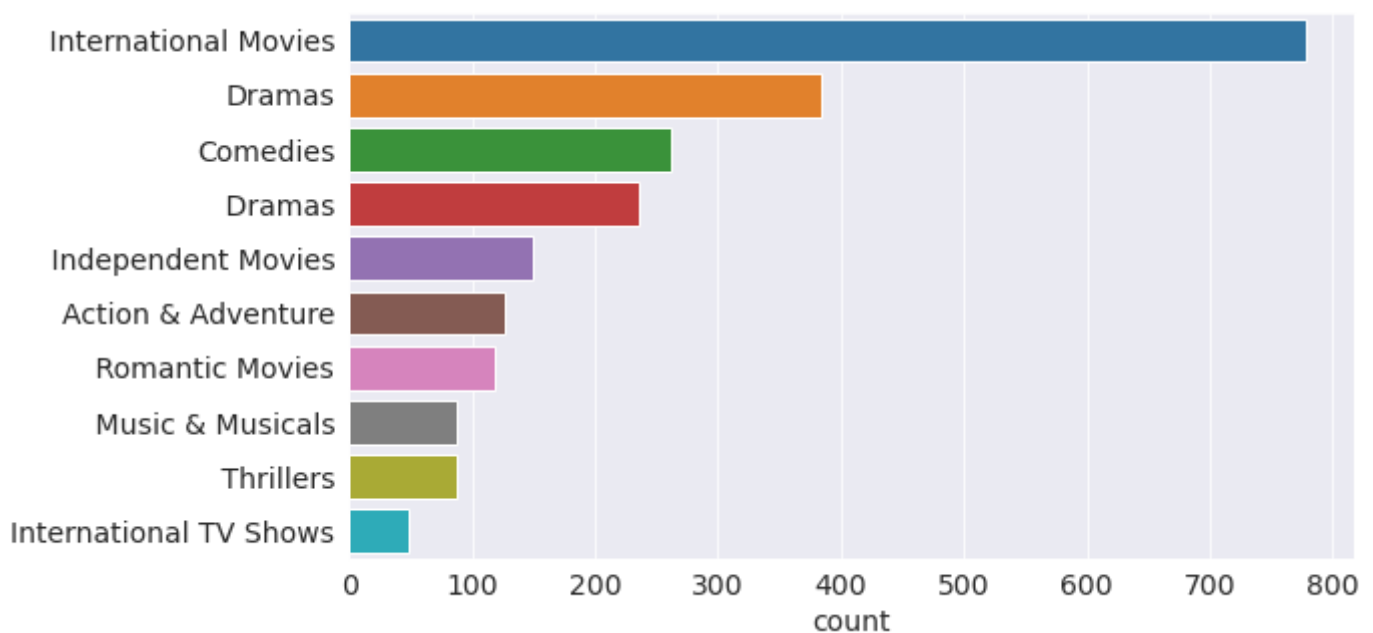
title

```

Kota Factory      International TV Shows
Kota Factory      Romantic TV Shows
Kota Factory      TV Comedies
Jeans             Comedies
Jeans             International Movies
...
Zenda             Dramas
Zenda             International Movies
Zubaan            Dramas
Zubaan            International Movies
Zubaan            Music & Musicals
Length: 2626, dtype: object

```

```
sns.countplot(y=Netflix_india_geners, order= Netflix_india_geners.value_counts().index[
```



According to dataset, International movies are most released in India on Netflix. and also Drama and comedies are on Netflix in India.

**Q4: In how many movies / Tv shows, Anupam Kher was cast ?**

```

Netflix_anupam = Netflix_df[Netflix_df['cast'].str.contains('Anupam Kher')]
Netflix_anupam.head(2)

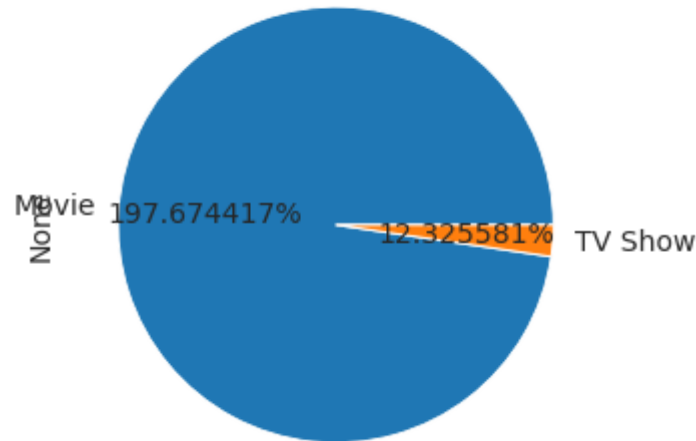
```

|     | show_id | type  | title             | director       | cast  | country | date_added | release_year | rating | duration | lis                |
|-----|---------|-------|-------------------|----------------|---|---------|------------|--------------|--------|----------|--------------------|
| 192 | s193    | Movie | C Kkompany        | Sachin Yardi   | Mithun Chakraborty, Tusshar Kapoor, Anupam Khe... | India   | 2021-08-27 | 2008         | TV-14  | 127 min  | Ac Adv Cor Interna |
| 202 | s203    | Movie | Kyaa Kool Hai Hum | Sangeeth Sivan | Tusshar Kapoor, Riteish Deshmukh, Isha Koppika... | India   | 2021-08-27 | 2005         | TV-MA  | 165 min  | Cor Interna M M    |

```
Netflix_anupam_df = Netflix_anupam.set_index('title').type.str.split(',', expand=True).s
```

```
Netflix_anupam_df.value_counts().plot(kind='pie', autopct= '1%2f%');  
plt.title('Contributor of Anupam kher on Netflix');
```

Contributor of Anupam kher on Netflix



In this pie chart shows, how many contibutor on Netflix,

- Anupam kher release movies on Netflix around = 197.67%
- Anupam kher release Tv show on Netflix around = 12.32% so Anupam kher is a highets popular, and Contributor on Netflix according to Datasets

## Q5 Which month has the most content add on Netlix ? show with Graph !

First off all, we will done seprate of month from Date\_added columns.

```
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 | 2021-09-25 | 2020         | PG-13  | 90 min    | Documentarie                                 |
| 1 | s2      | TV Show | Blood & Water        | No Director     | Ama Qamata, Khosi Ngema, Gail Mabalan, Thaban... | South Africa  | 2021-09-24 | 2021         | TV-MA  | 2 Seasons | International TV Shows, T Dramas, T Mysterie |

First off all, we will done seprate of month from Date\_added columns in Datasets

```
Netflix_df['month'] = Netflix_df['date_added'].dt.month_name()
```

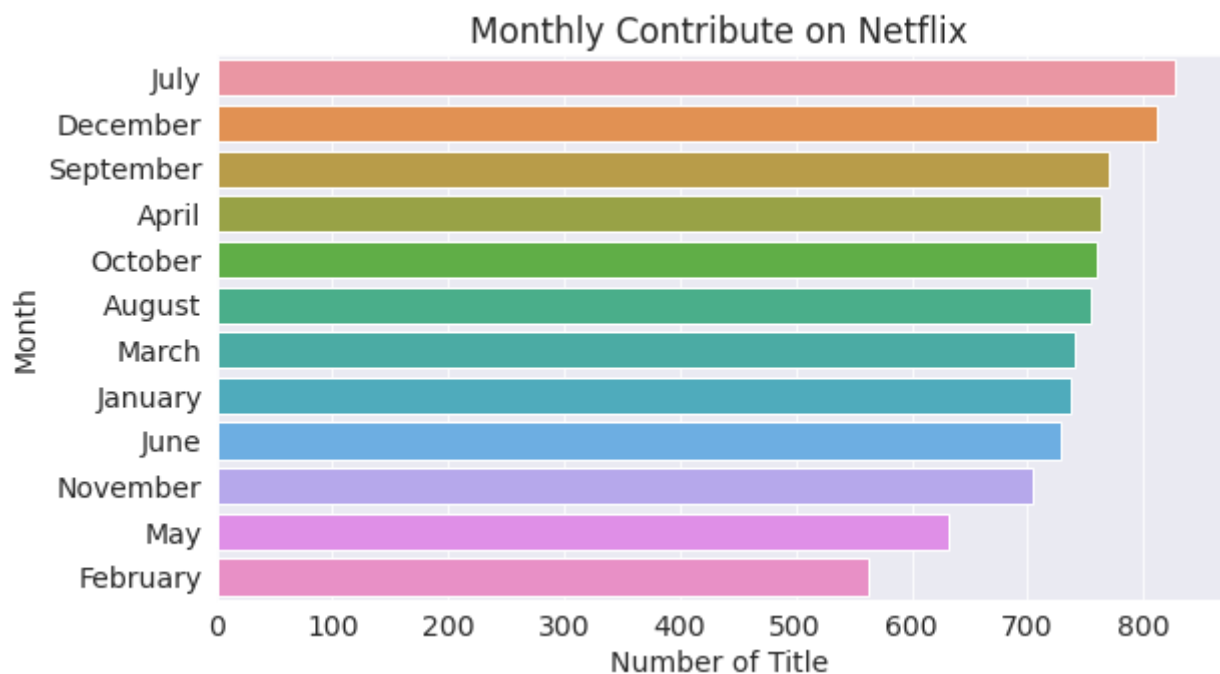
```
Netflix_df.head(2)
```

|   | 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 | 2021-09-25 | 2020         | PG-13  | 90 min    | Documentarie                                  |
| 1 | s2      | TV Show | Blood & Water        | No Director     | Ama Qamata, Khosi Ngema, Gail Mablane, Thaban... | South Africa  | 2021-09-24 | 2021         | TV-MA  | 2 Seasons | International TV Shows, T Dramas, T Mysteries |

```
Netflix_month = Netflix_df.set_index('title').month.str.split(',', expand=True).stack().Netflix_month
```

```
title
Dick Johnson Is Dead    September
Blood & Water           September
Ganglands               September
Jailbirds New Orleans   September
Kota Factory            September
...
Zodiac                  November
Zombie Dumb              July
Zombieland               November
Zoom                    January
Zubaan                   March
Length: 8793, dtype: object
```

```
sns.countplot(y = Netflix_month, order= Netflix_month.value_counts().index[:13])
plt.xlabel('Number of Title')
plt.ylabel('Month')
plt.title('Monthly Contribute on Netflix');
```



In all over Dataset, we founded, that at **July** and **December** months are mostly added content on Netflix

- Add on july around = 850
- and December around = 820
- and september , April and october are also high content added on Netflix

**Q6 Which top 3 movies / Tv shows on Netflix has TV-MA rating release in 2021 ? show with graph !**

```
Netflix_top_3 = Netflix_df[(Netflix_df['rating']=='TV-MA') & (Netflix_df['release_year']
Netflix_top_3.head(3)
```

| show_id | type         | title           | director  | cast  | country | date_added | release_year | rating | duration  |
|---------|--------------|-----------------|---|---|---------|------------|--------------|--------|-----------|
| 4       | s5 TV Show   | Kota Factory    | No Director                                       | Mayur More, Jitendra Kumar, Ranjan Raj, Alam K... | India   | 2021-09-24 | 2021         | TV-MA  | 2 Seasons |
| 297     | s298 TV Show | Navarasa        | Bejoy Nambiar, Priyadarshan, Karthik Narain, V... | Suriya, Vijay Sethupathi, Revathy, Prakash Raj... | India   | 2021-08-06 | 2021         | TV-MA  | 1 Season  |
| 551     | s552 Movie   | Haseen Dillruba | Vinil Mathew                                      | Taapsee Pannu, Vikrant Massey, Harshvardhan Ra... | India   | 2021-07-02 | 2021         | TV-MA  | 136 min   |

```
Netflix_top_3['title'].head(3)
```

```
4          Kota Factory
297         Navarasa
```



551 Haseen Dillruba  
Name: title, dtype: object

This are top 3 Movies/tv shows has TV-MA rating release in 2021,

- 2 Tv Shows- Kodta Factory, Navarasa
- 1 movie = Haseen Dillruba

Let us save and upload our work to Jovian before continuing.

```
import jovian
```

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

```
[jovian] Updating notebook "amodi1486/netflix-movies-and-tv-shows" on https://jovian.ai  
[jovian] Committed successfully! https://jovian.ai/amodi1486/netflix-movies-and-tv-shows
```

```
'https://jovian.ai/amodi1486/netflix-movies-and-tv-shows'
```

## Inferences and Conclusion

We have drawn many interesting inferences from the dataset Netflix titles; here's a summary of the few of them

- International Movies is a genre that is mostly in Netflix,
- movies is type that is mostly release on Netflix as compare to Tv shows,
- mostly movies and Tv shows release on 2018 and and after 2018 fall down release movise and tv show because od covid 19 and highest release movies and tv shows in india 2017, and after 2017 is fall down because of covid 19,
- The largest count of Netflix content is made with a "TV-MA" rating and TV-MA For Mature Audiences. May not be suitable for ages 17 and under and next is TA-14 rating high on Netflix Tv-14 Parents strongly cautioned. May not be suitable for ages 14 and under,
- The most popular actor on Netflix movie, based on the number of titles, is Anupam Kher. and next other like Rupa Bhimani, Takahiro sakurai and om puri are also popular on Netflix,
- The country by the amount of the produces content is the United States on Netflix and second India is also high contributor on Netflix,
- The most popular director on Netflix , with the most titles, is Rajiv Chilaka and next Raul Campos and Jan Suter are also highes movies and Tv shows are release on Netflix,
- highest content added on Netflix in July and December on all over year,

So we found all information about Netflix.

```
import jovian
```

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

# References and Future Work

- We can analyze the missing country, dates, cast, and directors etc.

Check out the following resources to learn more about the dataset and tools used in this notebook:

- Datasets sources = <https://www.kaggle.com/datasets>
- pandas uses tutorial = <https://www.w3schools.com/python/pandas/default.asp>
- matplotlib tutorial = [https://www.tutorialspoint.com/python\\_data\\_science/python\\_matplotlib.htm](https://www.tutorialspoint.com/python_data_science/python_matplotlib.htm)
- seaborn uses = <https://www.section.io/engineering-education/seaborn-tutorial/>
- opendatasets Python library: <https://github.com/JovianML/opendatasets>

```
import jovian
```

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

```
[jovian] Updating notebook "amodi1486/netflix-movies-and-tv-shows" on https://jovian.ai  
[jovian] Committed successfully! https://jovian.ai/amodi1486/netflix-movies-and-tv-shows
```

```
'https://jovian.ai/amodi1486/netflix-movies-and-tv-shows'
```