# 1 Netflix Movies and TV Shows Data Set

# 1.0.1 Import Required Libraries

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

# 1.0.2 Load(import) Data Using Pandas

```
[10]: data=pd.read_csv('netflix_titles.csv')
df=pd.DataFrame(data)
df
```

[10]:		show_id	type	title	director	\	
	0	<b>s</b> 1	Movie	Dick Johnson Is Dead	Kirsten Johnson		
	1	s2	TV Show	Blood & Water	NaN		
	2	s3	TV Show	Ganglands	Julien Leclercq		
	3	s4	TV Show	Jailbirds New Orleans	NaN		
	4	s5	TV Show	Kota Factory	NaN		
	8802		Movie	Zodiac	David Fincher		
	8803		TV Show	Zombie Dumb	NaN		
	8804		Movie	Zombieland	Ruben Fleischer		
	8805	s8806	Movie	Zoom	Peter Hewitt		
	8806	s8807	Movie	Zubaan	Mozez Singh		
					cast	country \	
	0				NaN Unite	d States	
	1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban South Africa					
	2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi NaN					
	3				NaN	NaN	
	4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K India					
		M . I . D	CC . I I . I				
	8802	Mark R	uttaio, Jak	ce Gyllenhaal, Robert Do	•		
	8803				NaN	NaN	

```
8804 Jesse Eisenberg, Woody Harrelson, Emma Stone, ... United States
8805 Tim Allen, Courteney Cox, Chevy Chase, Kate Ma... United States
8806 Vicky Kaushal, Sarah-Jane Dias, Raaghay Chanan...
                                                                India
                          release_year rating
             date added
                                                 duration \
0
     September 25, 2021
                                  2020
                                                   90 min
                                        PG-13
1
     September 24, 2021
                                  2021
                                        TV-MA 2 Seasons
2
     September 24, 2021
                                  2021
                                        TV-MA
                                                 1 Season
3
     September 24, 2021
                                  2021
                                        TV-MA
                                                1 Season
4
      September 24, 2021
                                  2021 TV-MA 2 Seasons
8802
       November 20, 2019
                                  2007
                                                  158 min
                                             R
           July 1, 2019
8803
                                               2 Seasons
                                  2018
                                        TV-Y7
8804
        November 1, 2019
                                  2009
                                                   88 min
                                             R
       January 11, 2020
                                            PG
8805
                                  2006
                                                   88 min
8806
          March 2, 2019
                                  2015 TV-14
                                                  111 min
                                               listed in
0
                                           Documentaries
1
        International TV Shows, TV Dramas, TV Mysteries
2
      Crime TV Shows, International TV Shows, TV Act...
3
                                  Docuseries, Reality TV
4
      International TV Shows, Romantic TV Shows, TV ...
                         Cult Movies, Dramas, Thrillers
8802
                 Kids' TV, Korean TV Shows, TV Comedies
8803
8804
                                Comedies, Horror Movies
8805
                     Children & Family Movies, Comedies
8806
         Dramas, International Movies, Music & Musicals
                                             description
0
      As her father nears the end of his life, filmm...
1
      After crossing paths at a party, a Cape Town t...
2
      To protect his family from a powerful drug lor...
      Feuds, flirtations and toilet talk go down amo...
3
4
      In a city of coaching centers known to train I...
8802 A political cartoonist, a crime reporter and a...
8803 While living alone in a spooky town, a young g...
8804 Looking to survive in a world taken over by zo...
8805 Dragged from civilian life, a former superhero...
8806 A scrappy but poor boy worms his way into a ty...
[8807 rows x 12 columns]
```

# 1.1 Data Cleaning

**Step-1**:Check is there any null or duplicated values is there or not

```
[16]: duplicate=df.duplicated().sum()
null=df.isnull().sum()
print("Duplicates:",duplicate)
print("Null Values:",null)
```

```
Duplicates: 0
Null Values: show_id
                                 0
type
title
                   0
director
                2634
cast
                 825
                 831
country
date_added
                  10
release_year
                   0
rating
                    4
                    3
duration
listed_in
                    0
description
                    0
dtype: int64
```

## So there is no duplicate values

And there are Null values in director column and cast column , country , date\_added column and rating , duration column

#### we don't know cast and director and country details so i will fill with unknown

```
[35]: df['director'].fillna('unknown')
    df['cast'].fillna('unknown')
    df['country'].fillna('unknown')
    null=df.isnull().sum() #to check the null values replaced or not
    print(null)
```

show_id	0
type	0
title	0
director	0
cast	0
country	0
date_added	10
release_year	0
rating	4
duration	3
listed_in	0
description	0
dtype: int64	

#### 1.1.1 Know Fill Remaining all with mode

```
[39]: df.fillna(df['date_added'].mode()[0],inplace=True) #it means it will fill with_
sfirst mode value

df.fillna(df['rating'].mode()[0],inplace=True)

df.fillna(df['duration'].mode()[0],inplace=True)

null=df.isnull().sum()

print(null)
```

show\_id 0 0 type title 0 0 director 0 cast 0 country date\_added 0 release\_year 0 rating 0 0 duration listed\_in 0 description 0 dtype: int64

#### 1.1.2 Cleaned the null and duplicates data

#### 1.1.3 Know Standardize text values like country names

```
[57]: df['title']=df['title'].str.strip().str.title()
df['country'] = df['country'].str.strip().str.title()
df.head(5)
```

```
title
[57]:
       show_id
                   type
                                                       director \
                          Dick Johnson Is Dead Kirsten Johnson
     0
            s 1
                  Movie
     1
            s2 TV Show
                                 Blood & Water
                                                       unknown
     2
            s3 TV Show
                                     Ganglands Julien Leclercq
     3
            s4 TV Show Jailbirds New Orleans
                                                       unknown
     4
            s5 TV Show
                                   Kota Factory
                                                       unknown
                                                                 country \
                                                     cast
     0
                                                  unknown United States
        Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban... South Africa
     1
     2
        Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...
                                                               Unknown
     3
                                                  unknown
                                                                 Unknown
        Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...
                                                                 India
                                                  duration \
                date_added
                             release_year rating
     O September 25, 2021
                                    2020 PG-13
                                                    90 min
     1 September 24, 2021
                                    2021 TV-MA 2 Seasons
```

2	September 24, 2021	2021	TV-MA	1 Season			
3	September 24, 2021	2021	TV-MA	1 Season			
4	September 24, 2021	2021	TV-MA	2 Seasons			
				listed_in	\		
0			Docu	umentaries	•		
1	International TV Shows, TV Dramas, TV Mysteries						
2	Crime TV Shows, International TV Shows, TV Act						
3		Docus	eries, R	Reality TV			
4	International TV Shows, Romantic TV Shows, TV						

## description

- 0 As her father nears the end of his life, filmm...
- 1 After crossing paths at a party, a Cape Town t...
- 2 To protect his family from a powerful drug lor...
- 3 Feuds, flirtations and toilet talk go down amo...
- 4 In a city of coaching centers known to train I...

# 1.1.4 Check the Data Types

# [64]: df.info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 8807 entries, 0 to 8806 Data columns (total 12 columns):

#	Column	Non-	Non-Null Count		
0	show_id	8807	non-null	object	
1	type	8807	non-null	object	
2	title	8807	non-null	object	
3	director	8807	non-null	object	
4	cast	8807	non-null	object	
5	country	8807	non-null	object	
6	date_added	8807	non-null	object	
7	release_year	8807	non-null	int64	
8	rating	8807	non-null	object	
9	duration	8807	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

#### 1.1.5 Convert date formats to a consistent type (e.g., dd-mm-yyyy)

```
[71]: df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')
      df.info()
      <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 8807 entries, 0 to 8806
     Data columns (total 12 columns):
                         Non-Null Count Dtype
          Column
      0
          show_id
                         8807 non-null
                                         object
      1
          type
                         8807 non-null
                                         object
      2
          title
                         8807 non-null
                                         object
      3
          director
                         8807 non-null
                                         object
      4
                                         object
          cast
                         8807 non-null
      5
          country
                         8807 non-null
                                         object
      6
          date_added
                         8719 non-null
                                         datetime64[ns]
      7
          release_year
                         8807 non-null
                                         int64
      8
          rating
                         8807 non-null
                                         object
      9
          duration
                                         object
                         8807 non-null
      10 listed_in
                         8807 non-null
                                         object
      11 description
                         8807 non-null
                                         object
     dtypes: datetime64[ns](1), int64(1), object(10)
     memory usage: 825.8+ KB
[73]: df.head(3)
[73]:
       show id
                     type
                                          title
                                                        director \
                   Movie Dick Johnson Is Dead Kirsten Johnson
      0
             s 1
      1
             s2 TV Show
                                 Blood & Water
                                                         unknown
      2
             s3 TV Show
                                     Ganglands Julien Leclercq
                                                       cast
                                                                   country \
      0
                                                    unknown United States
        Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban... South Africa
      2 Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...
                                                                 Unknown
       date_added
                     release_year rating
                                           duration \
      0 2021-09-25
                            2020 PG-13
                                             90 min
      1 2021-09-24
                            2021 TV-MA
                                          2 Seasons
      2 2021-09-24
                            2021 TV-MA
                                           1 Season
                                                  listed in
      0
                                              Documentaries
      1
           International TV Shows, TV Dramas, TV Mysteries
      2 Crime TV Shows, International TV Shows, TV Act...
```

#### description

- 0 As her father nears the end of his life, filmm...
- 1 After crossing paths at a party, a Cape Town t...
- 2 To protect his family from a powerful drug lor...

## **Date Column Changed Successfully Changed**

#### 1.1.6 Renaming column headers to be clean and uniform (e.g., lowercase, no spaces).

```
[82]: df.columns = df.columns.str.strip().str.lower().str.replace('_',")
      df.head(2)
[82]: showid
                                       title
                                                     director
                   type
                 Movie Dick Johnson Is Dead Kirsten Johnson
           s2 TV Show
                               Blood & Water
                                                      unknown
      1
                                                     cast
                                                                 country \
      0
                                                  unknown United States
      1 Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban... South Africa
        dateadded releaseyear rating
                                        duration \
      0 2021-09-25
                          2020 PG-13
                                          90 min
      1 2021-09-24
                          2021 TV-MA 2 Seasons
                                               listedin \
      0
                                          Documentaries
       International TV Shows, TV Dramas, TV Mysteries
                                              description
      0 As her father nears the end of his life, filmm...
      1 After crossing paths at a party, a Cape Town t...
```

[85]: df.info() #To Check data types

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
0	showid	8807 non-null	object
1	type	8807 non-null	object
2	title	8807 non-null	object
3	director	8807 non-null	object
4	cast	8807 non-null	object
5	country	8807 non-null	object
6	dateadded	8719 non-null	datetime64[ns]
7	releaseyear	8807 non-null	int64
8	rating	8807 non-null	object

9 duration 8807 non-null object 10 listedin 8807 non-null object 11 description 8807 non-null object dtypes: datetime64[ns](1), int64(1), object(10)

memory usage: 825.8+ KB

## 1.2 Data Cleaning Successfully Completed

[88]: #This is the Data After Cleaned df [88]: showid type title director \ Dick Johnson Is Dead Kirsten Johnson 0 s 1 Movie **Blood & Water** 1 s2 TV Show unknown 2 TV Show **s**3 Ganglands Julien Leclerca 3 Jailbirds New Orleans s4 TV Show unknown 4 **s**5 TV Show **Kota Factory** unknown David Fincher 8802 s8803 Movie Zodiac 8803 s8804 TV Show Zombie Dumb unknown Zombieland Ruben Fleischer 8804 s8805 Movie 8805 s8806 Movie Zoom Peter Hewitt 8806 s8807 Movie Zubaan Mozez Singh cast country \ 0 unknown United States 1 Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban... South Africa 2 Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi... Unknown 3 Unknown 4 Mayur More, Jitendra Kumar, Ranjan Raj, Alam K... India Mark Ruffalo, Jake Gyllenhaal, Robert Downey J... 8802 United States 8803 unknown Unknown 8804 Jesse Eisenberg, Woody Harrelson, Emma Stone, ... United States 8805 Tim Allen, Courteney Cox, Chevy Chase, Kate Ma... United States 8806 Vicky Kaushal, Sarah-Jane Dias, Raaghay Chanan... India dateadded releaseyear rating duration 90 min 0 2021-09-25 2020 PG-13 1 2021-09-24 2021 TV-MA 2 Seasons 2 2021-09-24 TV-MA 1 Season 2021 3 2021-09-24 2021 TV-MA 1 Season 4 2021 TV-MA 2 Seasons 2021-09-24 8802 2019-11-20 2007 R 158 min 2 Seasons 8803 2019-07-01 2018 TV-Y7 8804 2019-11-01 2009 R 88 min

8805 2020-01-11		2006	PG	88 min			
8806 2019-03-02		2015	TV-14	111 min			
	0 1 2 3 4	International Crime TV Shows International TV	s, Interna	tional T\ Doc	ımas, TV M / Shows, T useries, R	V Act eality TV	\
	8802 8803 8804 8805 8806	Kids Dramas, Inte	s' TV, Kor Children	ean TV S Come & Family	dies, Horr Movies, C	Comedies or Movies Comedies	
	0 1 2 3 4	As her father r After crossing r To protect his Feuds, flirtation In a city of coa	paths at family frons and t	a party, om a po oilet tall	his life, a Cape To werful dru < go dow	own t Ig lor n amo	
	8803 8804 8805	A political cart While living alo Looking to surv Dragged from c A scrappy but	ne in a s vive in a ivilian lit	spooky to world ta fe, a foo	own, a you ken over rmer supe	ung g by zo rhero	
	[8807 rows x 12 columns]						