

In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
df = pd.read_csv("C:/Users/ameya/OneDrive/Desktop/DSBDAL/NetFlix.csv")
```

In [4]:

```
df.isnull().sum()
```

Out[4]:

```
show_id      0
type         0
title        0
director    2389
cast        718
country     507
date_added   10
release_year  0
rating       7
duration     0
genres       0
description  0
dtype: int64
```

In [5]:

```
df.head()
```

Out[5]:

	show_id	type	title	director	cast	country	date_added	release_year	r
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	14-Aug-20	2020	
1	s10	Movie	1920	Vikram Bhatt	Rajneesh Duggal, Adah Sharma, Indraneil Sengup...	India	15-Dec-17	2008	
2	s100	Movie	3 Heroines	Iman Brotoseno	Reza Rahadian, Bunga Citra Lestari, Tara Basro...	Indonesia	05-Jan-19	2016	
3	s1000	Movie	Blue Mountain State: The Rise of Thadland	Lev L. Spiro	Alan Ritchson, Darin Brooks, James Cade, Rob R...	United States	01-Mar-16	2016	
4	s1001	TV Show	Blue Planet II	NaN	David Attenborough	United Kingdom	03-Dec-18	2017	



In [6]:

```
df.tail()
```

Out[6]:

	show_id	type	title	director	cast	country	date_added	release_year
7782	s995	TV Show	Blown Away	NaN	NaN	Canada	12-Jul-19	2019
7783	s996	TV Show	Blue Exorcist	NaN	Nobuhiko Okamoto, Jun Fukuyama, Kana Hanazawa,...	Japan	01-Sep-20	2017
7784	s997	Movie	Blue Is the Warmest Color	Abdellatif Kechiche	Léa Seydoux, Adèle Exarchopoulos, Salim Kechio...	France, Belgium, Spain	26-Aug-16	2013
7785	s998	Movie	Blue Jasmine	Woody Allen	Cate Blanchett, Sally Hawkins, Alec Baldwin, L...	United States	08-Mar-19	2013
7786	s999	Movie	Blue Jay	Alex Lehmann	Sarah Paulson, Mark Duplass, Clu Gulager	United States	06-Dec-16	2016



In [8]:

```
df.describe()
```

Out[8]:

	release_year	duration
count	7787.000000	7787.000000
mean	2013.932580	69.122769
std	8.757395	50.950743
min	1925.000000	1.000000
25%	2013.000000	2.000000
50%	2017.000000	88.000000
75%	2018.000000	106.000000
max	2021.000000	312.000000

In [11]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7787 entries, 0 to 7786
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   show_id         7787 non-null   object
 1   type            7787 non-null   object
 2   title           7787 non-null   object
 3   director        5398 non-null   object
 4   cast            7069 non-null   object
 5   country         7280 non-null   object
 6   date_added      7777 non-null   object
 7   release_year    7787 non-null   int64
 8   rating          7780 non-null   object
 9   duration        7787 non-null   int64
10   genres          7787 non-null   object
11   description     7787 non-null   object
dtypes: int64(2), object(10)
memory usage: 730.2+ KB
```

In [15]:

```
df.size
```

Out[15]:

93444

In [17]:

```
df.shape
```

Out[17]:

(7787, 12)

In [20]:

```
df.ndim
# returns 1 for 1d and 2 for 2d
```

Out[20]:

2

In [21]:

```
df.isnull().sum()
```

Out[21]:

```
show_id      0
type         0
title        0
director    2389
cast        718
country     507
date_added   10
release_year  0
rating       7
duration     0
genres       0
description  0
dtype: int64
```

In [22]:

```
df['rating'].value_counts()
```

Out[22]:

```
TV-MA      2863
TV-14      1931
TV-PG       806
R           665
PG-13       386
TV-Y        280
TV-Y7       271
PG          247
TV-G        194
NR           84
G           39
TV-Y7-FV     6
UR           5
NC-17        3
Name: rating, dtype: int64
```

In [23]:

```
df['rating'].fillna("TV-MA",inplace = True)
```

In [24]:

```
df.isnull().sum()
```

Out[24]:

```
show_id      0
type         0
title        0
director    2389
cast         718
country      507
date_added   10
release_year  0
rating       0
duration     0
genres       0
description  0
dtype: int64
```

In [25]:

```
df.dtypes
```

Out[25]:

```
show_id      object
type         object
title        object
director     object
cast         object
country      object
date_added   object
release_year  int64
rating       object
duration     int64
genres       object
description  object
dtype: object
```

In [26]:

```
duration_mean = df['duration'].mean()
```

In [27]:

```
duration_mean
```

Out[27]:

```
69.12276871709258
```

In [28]:

```
df['duration'].fillna(duration_mean,inplace = True)
```

In [29]:

```
df.isnull().sum()
```

Out[29]:

```
show_id      0
type         0
title        0
director    2389
cast        718
country     507
date_added   10
release_year  0
rating       0
duration     0
genres       0
description  0
dtype: int64
```

In [30]:

```
df['type'].value_counts()
```

Out[30]:

```
Movie      5377
TV Show    2410
Name: type, dtype: int64
```

In [31]:

```
df_type = pd.get_dummies(df['type'])
```

In [32]:

```
df_type
```

Out[32]:

	Movie	TV Show
0	0	1
1	1	0
2	1	0
3	1	0
4	0	1
...
7782	0	1
7783	0	1
7784	1	0
7785	1	0
7786	1	0

7787 rows × 2 columns

In [35]:

```
new_df = pd.concat([df,df_type],axis = 1)
```


In [37]:

```
new_df.head()
```

Out[37]:

	show_id	type	title	director	cast	country	date_added	release_year	r
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	14-Aug-20	2020	
1	s10	Movie	1920	Vikram Bhatt	Rajneesh Duggal, Adah Sharma, Indraneil Sengup...	India	15-Dec-17	2008	
2	s100	Movie	3 Heroines	Iman Brotoseno	Reza Rahadian, Bunga Citra Lestari, Tara Basro...	Indonesia	05-Jan-19	2016	
3	s1000	Movie	Blue Mountain State: The Rise of Thadland	Lev L. Spiro	Alan Ritchson, Darin Brooks, James Cade, Rob R...	United States	01-Mar-16	2016	
4	s1001	TV Show	Blue Planet II	NaN	David Attenborough	United Kingdom	03-Dec-18	2017	



In [42]:

```
new_df.drop(['type'],axis =1,inplace=True)
```

In [43]:

```
new_df.head()
```

Out[43]:

	show_id	title	director	cast	country	date_added	release_year	rating	c
0	s1	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	14-Aug-20	2020	TV-MA	
1	s10	1920	Vikram Bhatt	Rajneesh Duggal, Adah Sharma, Indraneil Sengup...	India	15-Dec-17	2008	TV-MA	
2	s100	3 Heroines	Iman Brotoseno	Reza Rahadian, Bunga Citra Lestari, Tara Basro...	Indonesia	05-Jan-19	2016	TV-PG	
3	s1000	Blue Mountain State: The Rise of Thadland	Lev L. Spiro	Alan Ritchson, Darin Brooks, James Cade, Rob R...	United States	01-Mar-16	2016	R	
4	s1001	Blue Planet II	NaN	David Attenborough	United Kingdom	03-Dec-18	2017	TV-G	



In []: