

```
In [1]: import pandas as pd
import numpy as ny
```

```
In [2]: df= pd.read_csv("NetFlix.csv")
df.head()
```

Out[2]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	genres	description
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	14-Aug-20	2020	TV-MA	4	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...
1	s10	Movie	1920	Vikram Bhatt	Rajneesh Duggal, Adah Sharma, Indraneil Sengup...	India	15-Dec-17	2008	TV-MA	143	Horror Movies, International Movies, Thrillers	An architect and his wife move into a castle t...
2	s100	Movie	3 Heroines	Iman Brotoseno	Reza Rahadian, Bunga Citra Lestari, Tara Basro...	Indonesia	05-Jan-19	2016	TV-PG	124	Dramas, International Movies, Sports Movies	Three Indonesian women break records by becomi...
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	R	90	Comedies	New NFL star Thad buys his old teammates' belo...
4	s1001	TV Show	Blue Planet II	NaN	David Attenborough	United Kingdom	03-Dec-18	2017	TV-G	1	British TV Shows, Docuseries, Science & Nature TV	This sequel to the award-winning nature series...

```
In [3]: 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 [4]: df.describe()
```

Out[4]:

	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 [8]: ny.max(df.type)
```

Out[8]: 'TV Show'

```
In [10]: ny.std(df.duration)
```

Out[10]: 50.94747171259882

```
In [11]: ny.var(df.duration)
```

Out[11]: 2595.6448739060565

```
In [14]: ny.mean(df.duration)
```

```
Out[14]: 69.12276871709258
```

```
In [15]: df.isnull().sum()
```

```
Out[15]: 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 [16]: 1 pd.isna(df.director)
```

```
Out[16]: 0      True
         1     False
         2     False
         3     False
         4      True
         ...
        7782    True
        7783    True
        7784    False
        7785    False
        7786    False
         Name: director, Length: 7787, dtype: bool
```

```
In [17]: df.dtypes
```

```
Out[17]: 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 [18]: size=df.size
         format(size)
```

```
Out[18]: '93444'
```

```
In [19]: shape = df.shape
         format(shape)
```

```
Out[19]: '(7787, 12)'
```

```
In [23]: dim=df.ndim
         print("ndim of dataframe ")
         format(dim)
```

```
ndim of dataframe
```

```
Out[23]: '2'
```

```
In [24]: type(df.show_id[0])
```

```
Out[24]: str
```

```
In [ ]: // qualitative data converted into quantative data.
```

```
In [25]: df_type = pd.get_dummies(df['type'])
```

```
In [26]: df_type
```

Out[26]:

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

```
In [29]: new_df
```

Out[29]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	genres	description	Movie	TV Show
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	14-Aug-20	2020	TV-MA	4	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...	0	1
1	s10	Movie	1920	Vikram Bhatt	Rajneesh Duggal, Adah Sharma, Indraneil Sengup...	India	15-Dec-17	2008	TV-MA	143	Horror Movies, International Movies, Thrillers	An architect and his wife move into a castle t...	1	0
2	s100	Movie	3 Heroines	Iman Brotoseno	Reza Rahadian, Bunga Citra Lestari, Tara Basro...	Indonesia	05-Jan-19	2016	TV-PG	124	Dramas, International Movies, Sports Movies	Three Indonesian women break records by becomi...	1	0
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	R	90	Comedies	New NFL star Thad buys his old teammates' belo...	1	0
4	s1001	TV Show	Blue Planet II	NaN	David Attenborough	United Kingdom	03-Dec-18	2017	TV-G	1	British TV Shows, Docuseries, Science & Nature TV	This sequel to the award-winning nature series...	0	1
...
7782	s995	TV Show	Blown Away	NaN	NaN	Canada	12-Jul-19	2019	TV-14	1	International TV Shows, Reality TV	Ten master artists turn up the heat in glassbl...	0	1
7783	s996	TV Show	Blue Exorcist	NaN	Nobuhiko Okamoto, Jun Fukuyama, Kana Hanazawa,...	Japan	01-Sep-20	2017	TV-MA	2	Anime Series, International TV Shows	Determined to throw off the curse of being Sat...	0	1
7784	s997	Movie	Blue Is the Warmest Color	Abdellatif Kechiche	Léa Seydoux, Adèle Exarchopoulos, Salim Kechio...	France, Belgium, Spain	26-Aug-16	2013	NC-17	180	Dramas, Independent Movies, International Movies	Determined to fall in love, 15-year-old Adele ...	1	0
7785	s998	Movie	Blue Jasmine	Woody Allen	Cate Blanchett, Sally Hawkins, Alec Baldwin, L...	United States	08-Mar-19	2013	PG-13	98	Comedies, Dramas, Independent Movies	The high life leads to high anxiety for a fash...	1	0
7786	s999	Movie	Blue Jay	Alex Lehmann	Sarah Paulson, Mark Duplass, Clu Gulager	United States	06-Dec-16	2016	TV-MA	81	Dramas, Independent Movies, Romantic Movies	Two former high school sweethearts unexpectedl...	1	0

7787 rows × 14 columns

```
In [30]: df['type'].map(type) == str
```

```
Out[30]: 0      True
         1      True
         2      True
         3      True
         4      True
         ...
        7782    True
        7783    True
        7784    True
        7785    True
        7786    True
        Name: type, Length: 7787, dtype: bool
```

```
In [31]: (df['type'].map(type) == str).sum()
```

```
Out[31]: 7787
```

```
In [32]: (df['type'].map(type) == str).sum() == len(df)
```

```
Out[32]: True
```

```
In [33]: df['country'].map(type) == str
```

```
Out[33]: 0      True
         1      True
         2      True
         3      True
         4      True
         ...
        7782    True
        7783    True
        7784    True
        7785    True
        7786    True
        Name: country, Length: 7787, dtype: bool
```

```
In [35]: (df['duration'].map(type) == str).sum() == len(df)
```

```
Out[35]: False
```

```
In [36]: (df['duration'].map(type) == str).sum()
```

```
Out[36]: 0
```

```
In [37]: len(df) - (df['duration'].map(type) == str).sum()
```

```
Out[37]: 7787
```

```
In [38]: for i in df.duration:
         if(type(i)!=str):
             print(i)
```

```
100
105
1
2
5
81
105
1
116
77
1
61
110
110
103
93
94
1
51
118
-
```

```
In [39]: len(df) - (df['release_year'].map(type) == str).sum()
```

```
Out[39]: 7787
```

```
In [40]: df['country'].map(type)
```

```
Out[40]: 0      <class 'str'>
         1      <class 'str'>
         2      <class 'str'>
         3      <class 'str'>
         4      <class 'str'>
         ...
        7782    <class 'str'>
        7783    <class 'str'>
        7784    <class 'str'>
        7785    <class 'str'>
        7786    <class 'str'>
        Name: country, Length: 7787, dtype: object
```

```
In [41]: type(df.country[2])
```

```
Out[41]: str
```

```
In [43]: pd.isna(df.country).sum()
```

```
Out[43]: 507
```

```
In [ ]:
```