```
import pandas as pd
import numpy as np

import plotly.express as px
import plotly.graph_objects as go
import plotly.figure_factory as ff
from plotly.subplots import make_subplots

import matplotlib.pyplot as plt

import seaborn as sns


df = pd.read_csv("/content/netflix_titles.csv")
df.head()
```

	show_id	type	title	director	cast	country	date_ad
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	Augus
1	s2	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	Decer 23,
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	Decer 20,
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	Nover 16,
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	Janua

Understanding Dataset

```
df.shape
```

```
(7787, 12)
```

```
df.duplicated().any()
```

```
False
```

```
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   object
 10  listed_in     7787 non-null   object
 11  description   7787 non-null   object
dtypes: int64(1), object(11)
memory usage: 730.2+ KB
```

```
df.nunique()
```

```
show_id         7787
type               2
title           7787
director        4049
cast            6831
country          681
date_added      1565
release_year      73
rating            14
duration         216
listed_in        492
description     7769
dtype: int64
```

```
df.nunique() / df.shape[0] * 100
```

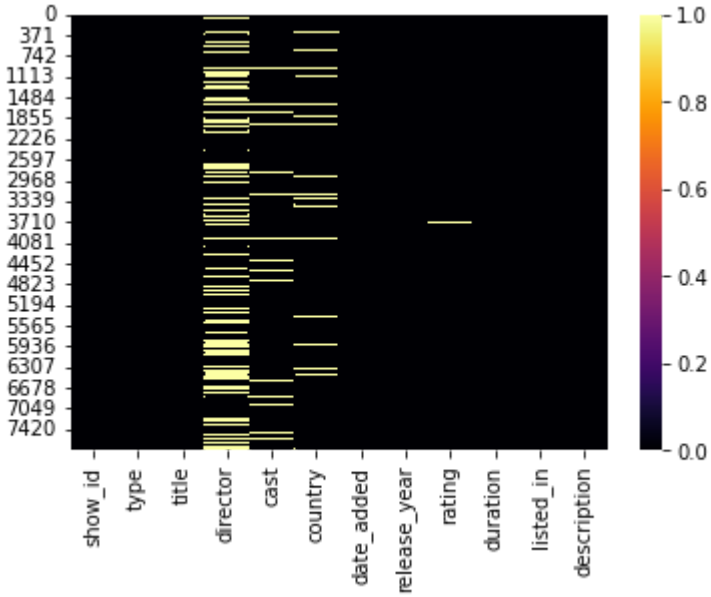
```
show_id         100.000000
type              0.025684
title           100.000000
director         51.996918
cast             87.723128
country           8.745345
date_added       20.097599
release_year      0.937460
rating            0.179787
duration          2.773854
listed_in         6.318223
description      99.768846
dtype: float64
```

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

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

```
sns.heatmap(df.isnull(),cmap = 'inferno')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fdf50390390>
```




```
df.isnull().sum() / df.shape[0] * 100
```

```
show_id          0.000000
type             0.000000
title            0.000000
director        30.679337
cast             9.220496
country          6.510851
date_added       0.128419
release_year     0.000000
rating           0.089893
duration         0.000000
listed_in        0.000000
description      0.000000
dtype: float64
```

Handling missing values

```
df.drop(['director','cast'], axis=1 ,inplace=True)
df.head()
```

	show_id	type	title	country	date_added	release_year	rating	dura
0	s1	TV Show	3%	Brazil	August 14, 2020	2020	TV-MA	4 Sea
1	s2	Movie	7:19	Mexico	December 23, 2016	2016	TV-MA	9
2	s3	Movie	23:59	Singapore	December 20, 2018	2011	R	7
3	s4	Movie	9	United States	November 16, 2017	2009	PG-13	8
4	s5	Movie	21	United States	January 1, 2020	2008	PG-13	12

```
df['country'].fillna('United States', inplace=True)
df['country'].isnull().any()
```

```
False
```

```
df.dropna(subset=['date_added'],axis = 0, inplace = True)
df['date_added'].isnull().any()
```

```
False
```

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

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

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

	show_id	type	title	country	date_added	release_year	rating	dur
67	s68	Movie	13TH: A Conversation with Oprah Winfrey & Ava ...	United States	January 26, 2017	2017	NaN	
2359	s2360	TV Show	Gargantia on the Verdurous Planet	Japan	December 1, 2016	2013	NaN	1 S
3660	s3661	TV Show	Little Lunch	Australia	February 1, 2018	2015	NaN	1 S
3736	s3737	Movie	Louis C.K. 2017	United States	April 4, 2017	2017	NaN	

Replacing the missing 'rating' values

			Louis C.K.:	United	September			

```
rating_replacements = {67: 'TV-PG', 2359: 'TV-14', 3660: 'TV-MA', 3736: 'TV-MA', 37
for index, rating in rating_replacements.items():
    df.loc[index, 'rating'] = rating
df['rating'].isnull().any()
```

```
False
```

Colab paid products - Cancel contracts here

✓ 0s completed at 23:35