

```
import numpy as np

import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv("netflix.csv")

data.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
```

```
data.head()
```

	show_id	type	title	director
0	s1	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

	cast	country
0	NaN	United 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

	date_added	release_year	rating	duration
0	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                                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 ...
```

```

                                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...
```

```
data.columns
```

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

```
data.ndim
```

```
2
```

```
data.describe()
```

```

      release_year
count  8807.000000
mean   2014.180198
std     8.819312
min    1925.000000
25%    2013.000000
50%    2017.000000
75%    2019.000000
max     2021.000000
```

```
data.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

```
print('\nColumns with missing value:')
```

```
print(data.isnull().any())
```

Columns with missing value:

show_id	False
---------	-------

type	False
------	-------

title	False
-------	-------

director	True
----------	------

cast	True
------	------

country	True
---------	------

date_added	True
------------	------

release_year	False
--------------	-------

rating	True
--------	------

duration	True
----------	------

listed_in	False
-----------	-------

description	False
-------------	-------

dtype: bool

```
data.T.apply(lambda x: x.isnull().sum(), axis = 1)
```

show_id	0
---------	---

type	0
------	---

title	0
-------	---

director	2634
----------	------

cast	825
------	-----

country	831
---------	-----

date_added	10
------------	----

release_year	0
--------------	---

rating	4
--------	---

duration	3
----------	---

listed_in	0
-----------	---

description	0
-------------	---

dtype: int64

```
data.isnull().sum().sum()
```

4307

```

data.director.fillna("No Director", inplace=True)
data.cast.fillna("No Cast", inplace=True)
data.country.fillna("Country Unavailable", inplace=True)
data.dropna(subset=["date_added", "rating"], inplace=True)
#data.duration.fillna("No duration", inplace=True)

```

```
data.isnull().any()
```

```

show_id      False
type         False
title        False
director     False
cast         False
country      False
date_added   False
release_year False
rating       False
duration     True
listed_in    False
description  False
dtype: bool

```

```
data.describe()
```

```

count    release_year
count    8793.000000
mean     2014.183100
std       8.824128
min       1925.000000
25%       2013.000000
50%       2017.000000
75%       2019.000000
max       2021.000000

```

```
data.head()
```

	show_id	type	title	director \
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson
1	s2	TV Show	Blood & Water	No Director
2	s3	TV Show	Ganglands	Julien Leclercq
3	s4	TV Show	Jailbirds New Orleans	No Director
4	s5	TV Show	Kota Factory	No Director

	country \	cast
0	United States	No Cast
1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Country Unavailable

```

3                                     No Cast  Country
Unavailable
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...
India

```

```

      date_added  release_year  rating  duration \
0  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                                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 ...

```

```

                                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...

```

```
data.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 8793 entries, 0 to 8806
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   show_id         8793 non-null   object
1   type            8793 non-null   object
2   title           8793 non-null   object
3   director        8793 non-null   object
4   cast            8793 non-null   object
5   country         8793 non-null   object
6   date_added      8793 non-null   object
7   release_year    8793 non-null   int64
8   rating          8793 non-null   object
9   duration        8790 non-null   object
10  listed_in       8793 non-null   object
11  description      8793 non-null   object

```

```
dtypes: int64(1), object(11)
```

```
memory usage: 893.0+ KB
```

```
data.describe()
```

	release_year
count	8793.000000
mean	2014.183100
std	8.824128
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

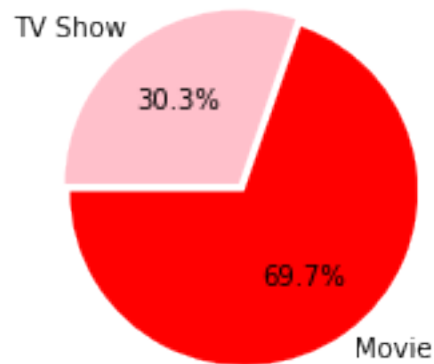
```
data.T.apply(lambda x: x.isnull().sum(), axis=1)
```

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

```
plt.figure(figsize=(6,3))
plt.title("Percontation of Netflix Titles that are either Movies or TV Shows")
g=plt.pie(data.type.value_counts(),explode=(0.025,0.025),
labels=data.type.value_counts().index,
colors=['red','pink'],autopct='%1.1f%%',
startangle=180)
```

Percentation of Netflix Titles that are either Movies or TV Shows



```
data["year_added"] = pd.to_datetime(data.date_added).dt.year
data_movies["year_added"] = pd.to_datetime(data_movies.date_added).dt.year
data_shows["year_added"] = pd.to_datetime(data_shows.date_added).dt.year
data_year = data.year_added.value_counts().to_frame().reset_index().rename(columns={"index": "year", "year_added": "count"})
data_year = data_year[data_year.year != 2020]
print(data_year)
```

```
filtered_genres = data.set_index('title').listed_in.str.split(', ', expand=True).stack().reset_index(level=1, drop=True);
plt.figure(figsize=(4,5))
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()
```

```
-----
-----
KeyError                                Traceback (most recent call last)
Input In [30], in <cell line: 1>()
----> 1 filtered_genres =
data.set_index('title').listed_in.str.split(', ',
expand=True).stack().reset_index(level=1, drop=True);
      2 plt.figure(figsize=(4,5))
      3 g = sns.countplot(y = filtered_genres,
      4 order=filtered_genres.value_counts().index[:20])

File ~\anaconda3\lib\site-packages\pandas\util\_decorators.py:311, in
deprecate_nonkeyword_arguments.<locals>.decorate.<locals>.wrapper(*arg
s, **kwargs)
```

```
305 if len(args) > num_allow_args:
306     warnings.warn(
307         msg.format(arguments=arguments),
308         FutureWarning,
309         stacklevel=stacklevel,
310     )
--> 311 return func(*args, **kwargs)

File ~\anaconda3\lib\site-packages\pandas\core\frame.py:5494, in
DataFrame.set_index(self, keys, drop, append, inplace,
verify_integrity)
5491         missing.append(col)
5493 if missing:
-> 5494     raise KeyError(f"None of {missing} are in the columns")
5496 if inplace:
5497     frame = self

KeyError: "None of ['title'] are in the columns"
```