

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

```
path = "/content/drive/MyDrive/netflix.csv"
data = pd.read_csv(path)
data.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA



```
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
```

```
movies = data.loc[df['type']=='Movie']
movies
```

```

    show_id  type      title      director      cast      country      date_added  release_
0          s1  Movie    Dick Johnson Is Dead    Kirsten Johnson      NaN      United States  September 25, 2021
6          s7  Movie    My Little Pony: A New Generation    Robert Cullen, José Luis Ucha    Vanessa Hudgens, Kimiko Glenn, James Marsden, ...      NaN  September 24, 2021
7          s8  Movie      Sankofa      Haile Gerima    Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...    United States, Ghana, Burkina Faso, United Kin...  September 24, 2021
9          s10  Movie    The Starling      Theodore Melfi    Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...      United States  September 24, 2021
...          ...      ...      ...      ...      ...      ...      ...      ...
tvshows = data.loc[df['type']=='TV Show']
tvshows
```

	show_id	type	title	director	cast	country	date_added	release_year
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021
...
8795	s8796	TV Show	Yu-Gi-Oh! Arc-V	NaN	Mike Liscio, Emily Bauer, Billy Bob Thompson, ...	Japan, Canada	May 1, 2018	2015
8796	s8797	TV Show	Yunus Emre	NaN	Gökhan Atalay, Payidar Tüfekçioğlu, Baran Akbu...	Turkey	January 17, 2017	2016
8797	s8798	TV Show	Zak Storm	NaN	Michael Johnston, Jessica Gee-George, Christin...	United States, France, South Korea, Indonesia	September 13, 2018	2016
8800	s8801	TV Show	Zindagi Gulzar Hai	NaN	Sanam Saeed, Fawad Khan, Ayesha Omer, Mehreen ...	Pakistan	December 15, 2016	2012
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018
2676 rows × 12 columns								



```

data['type'].value_counts()['Movie']

6131

data['type'].value_counts()['TV Show']

2676
```

There are more movies as compared to TV shows

```

data['rating'].unique()

array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
      'TV-G', 'G', 'NC-17', '74 min', '84 min', '66 min', 'NR', nan,
      'TV-Y7-FV', 'UR'], dtype=object)

data['rating'].value_counts()

TV-MA      3207
TV-14      2160
TV-PG       863
```

```

R          799
PG-13      490
TV-Y7      334
TV-Y       307
PG         287
TV-G       220
NR         80
G          41
TV-Y7-FV   6
NC-17      3
UR         3
74 min     1
84 min     1
66 min     1
Name: rating, dtype: int64

```

This shows that movies and tvshows with rating TV-MA are more liked by people.

```

movies['rating'].value_counts()['TV-MA']

2062

```

```

tvshows['rating'].value_counts()['TV-MA']

1145

```

This shows that movies with rating TV-MA are mre liked and watched by people rather than Tv shows with the same rating.

```

movies_per_year = movies.groupby('release_year').size()
movies_per_year.tail(20)

release_year
2002      44
2003      51
2004      55
2005      67
2006      82
2007      74
2008     113
2009     118
2010     154
2011     145
2012     173
2013     225
2014     264
2015     398
2016     658
2017     767
2018     767
2019     633
2020     517
2021     277
dtype: int64

```

This shows the number of movies released over past two decades. The year 2017 and 2018 has the most number of movie release.

```

tvshows_per_year = tvshows.groupby('release_year').size()
tvshows_per_year.tail(20)

release_year
2002      7
2003     10
2004      9
2005     13
2006     14
2007     14
2008     23
2009     34
2010     40
2011     40
2012     64
2013     63
2014     88
2015    162
2016    244
2017    265
2018    380
2019    397
2020    436
2021    315
dtype: int64

```

This shows that number of movie release over the past two decades and year 2020 has the most number of tvshows release followed by year 2019 and 2018.

This also shows that netflix is more focused on movies than tv shows.

```
movies['director'].value_counts()
```

```
Rajiv Chilaka          19
Raúl Campos, Jan Suter  18
Suhas Kadav            16
Marcus Raboy           15
Jay Karas              14
..
Dennis Rovira van Boekholt  1
Naoto Amazutsumi         1
Jenny Gage               1
Kaila York               1
Mozes Singh              1
Name: director, Length: 4354, dtype: int64
```

```
movies['cast'].value_counts()
```

```
Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jigna Bhardwaj, Rajesh Kava, Mousam, Swapnil
13
Samuel West
10
Jeff Dunham
7
Kevin Hart
6
Craig Sechler
6
..
Hánssel Casillas, Loreto Peralta, Andrea Sutton, Luis De La Rosa, Alejandro Flores, Anajosé Aldrete, Daniel Haddad, Montserrat
Marañón, Juan Carlos Viana Prieto, Hernán Del Riego      1
Oscar Martínez, Inma Cuesta, Mafalda Carbonell, Nacho López, Isabel Requena, Aina Clotet, Antonio Valero, María Zamora, Hugo
Balaguer, Valeria Schoneveld                            1
Casey Affleck, Rooney Mara, Liz Franke, Rob Zabrecky, Will Oldham, Sonia Acevedo
1
Faizon Love, Kirk Fox, Tony Todd, Pearl Thusi, Cassie Clare, Fiona Ramsey, Neels Clasen, Roxy Nel
1
Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary, Meghna Malik, Malkeet Rauni, Anita Shabdish, Chittaranjan
Tripathy                                                  1
Name: cast, Length: 5445, dtype: int64
```

```
tvshows['director'].value_counts()
```

```
Alastair Fothergill      3
Rob Seidenglanz          2
Hsu Fu-chun              2
Iginio Straffi            2
Shin Won-ho              2
..
Juliana Vicente          1
Chang Chin-jung, Chen Rong-hui  1
Thierry Demaizière, Alban Teurlai  1
Manolo Caro              1
Michael Cumming          1
Name: director, Length: 222, dtype: int64
```

```
tvshows['cast'].value_counts()
```

```
David Attenborough
14
Michela Luci, Jamie Watson, Anna Claire Bartlam, Dante Zee, Eric Peterson
4
Dave Chappelle
3
You, Reina Triendl, Yoshimi Tokui, Azusa Babazono, Ryota Yamasato, Shono Hayama
2
Chris Packham
2
..
Vir Das, Ranvir Shorey, Manoj Pahwa, Ravi Kishan, Raza Murad, Inaamulhaq, Suhail Nayyar, Amrita Bagchi
1
Sebastián Rulli, Renata Notni, Roberto Mateos, Irina Baeva, Cassandra Sánchez Navarro, Manuel Balbi, Javier Gómez, Alejandro
Ávila, Sofia Castro, Juan Pablo Gil      1
Kenya Barris, Rashida Jones, Iman Benson, Genneya Walton, Scarlet Spencer, Justin Claiborne, Ravi Cabot-Conyers
1
Lee Min-ho, Kim Go-eun, Woo Do-hwan, Kim Kyung-nam, Jung Eun-chae, Lee Jung-jin, Kim Young-ok
1
```

```

Sanam Saeed, Fawad Khan, Ayesha Omer, Mehreen Raheel, Sheheryar Munawar, Samina Peerzada, Waseem Abbas, Javed Sheikh, Hina Khawaja
Bayat
1
Name: cast, Length: 2283, dtype: int64

```

```
movies['country'].value_counts()
```

```

United States      2058
India              893
United Kingdom     206
Canada            122
Spain              97
...
United Kingdom, Russia, United States    1
Paraguay, Argentina                     1
United Kingdom, Malawi                   1
Austria, Iraq, United States             1
United Arab Emirates, Jordan              1
Name: country, Length: 651, dtype: int64

```

```
tvshows['country'].value_counts()
```

```

United States      760
United Kingdom     213
Japan              169
South Korea        158
India              79
...
Belarus            1
United Kingdom, Australia    1
France, Australia, Germany   1
Australia, New Zealand, United States    1
United States, France, South Korea, Indonesia    1
Name: country, Length: 196, dtype: int64

```

Comparing these data, it clearly shows that United States has access to most of the movies, followed by India. As for tvshows, United states has the access to most of the tvshows, followed by United Kingdom.

```
movies['date_added'].value_counts()
```

```

January 1, 2020      97
November 1, 2019     75
March 1, 2018        72
December 31, 2019    67
October 1, 2018      64
..
October 28, 2017     1
October 27, 2017     1
October 26, 2017     1
October 24, 2017     1
January 11, 2020     1
Name: date_added, Length: 1533, dtype: int64

```

```
tvshows['date_added'].value_counts()
```

```

July 6, 2021         31
January 1, 2016      23
August 1, 2017       22
June 19, 2021        21
July 1, 2017         19
..
July 20, 2019        1
July 18, 2019        1
July 6, 2019         1
July 5, 2019         1
September 13, 2018   1
Name: date_added, Length: 1051, dtype: int64

```

```
movies['duration'].value_counts()
```

```

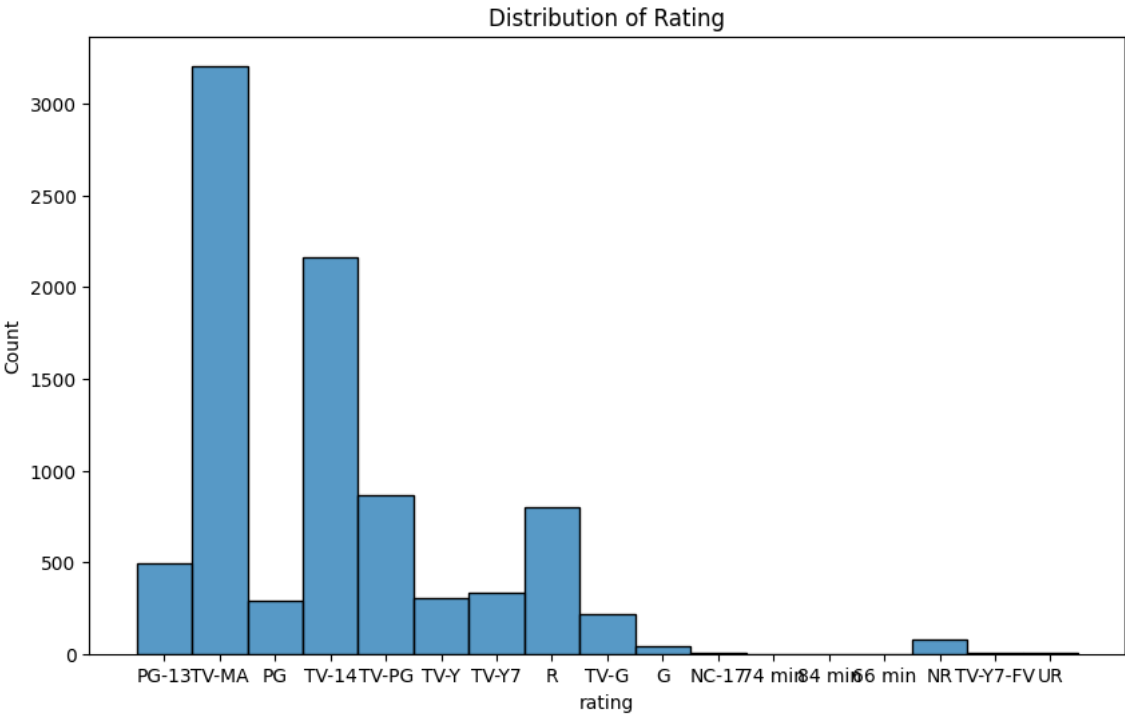
90 min      152
94 min      146
93 min      146
97 min      146
91 min      144
...
212 min      1
8 min        1
186 min      1
193 min      1
191 min      1
Name: duration, Length: 205, dtype: int64

```

```
movies['listed_in'].value_counts()

Dramas, International Movies      362
Documentaries                    359
Stand-Up Comedy                  334
Comedies, Dramas, International Movies  274
Dramas, Independent Movies, International Movies  252
...
Sci-Fi & Fantasy                  1
Sports Movies                    1
Children & Family Movies, Comedies, Cult Movies  1
Cult Movies, Dramas, Music & Musicals  1
Cult Movies, Dramas, Thrillers    1
Name: listed_in, Length: 278, dtype: int64
```

```
plt.figure(figsize=(10, 6))
sns.histplot(data["rating"])
plt.title("Distribution of Rating")
plt.show()
```

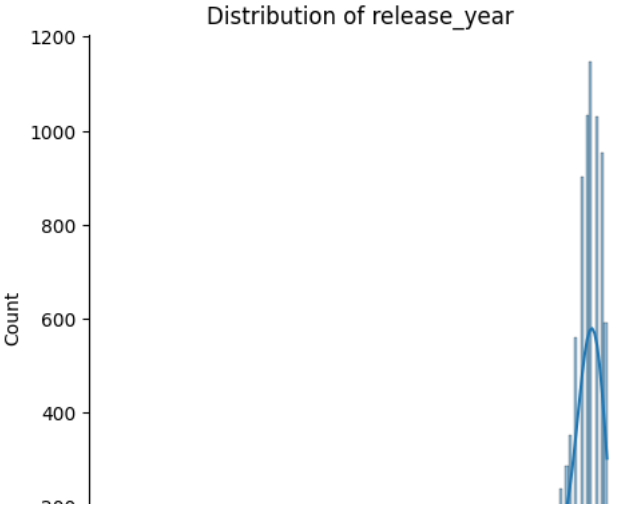
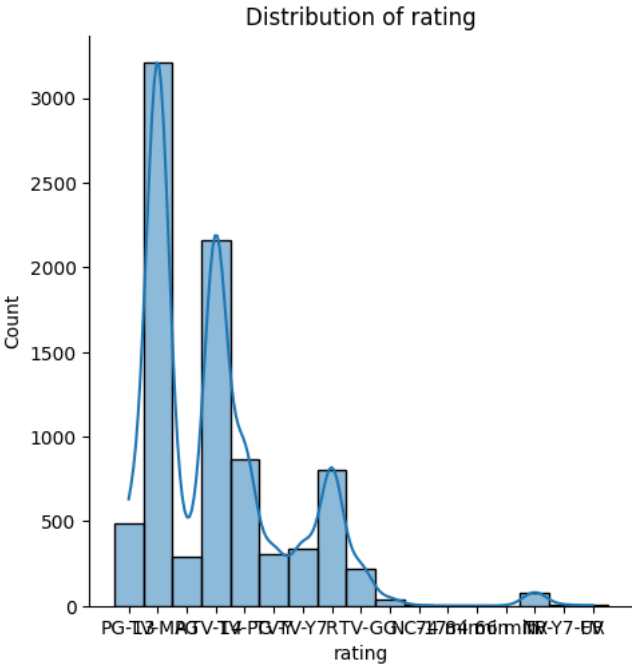
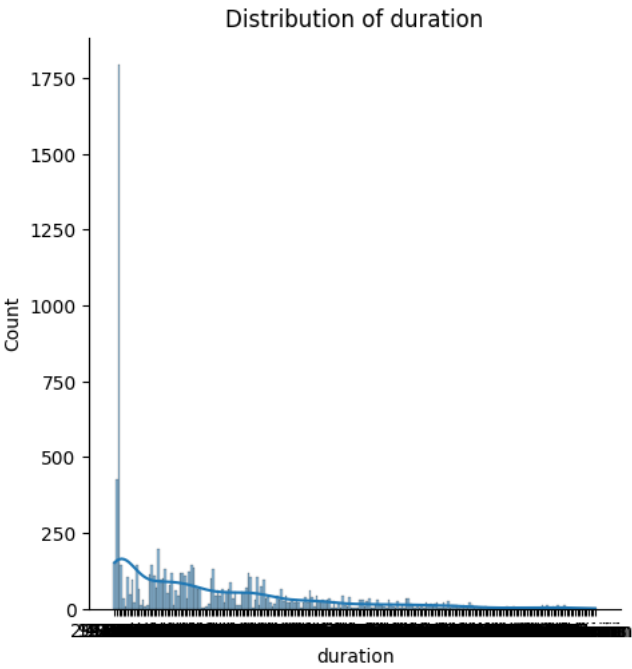


```
missing_values = data.isnull().sum()
print(missing_values)
```

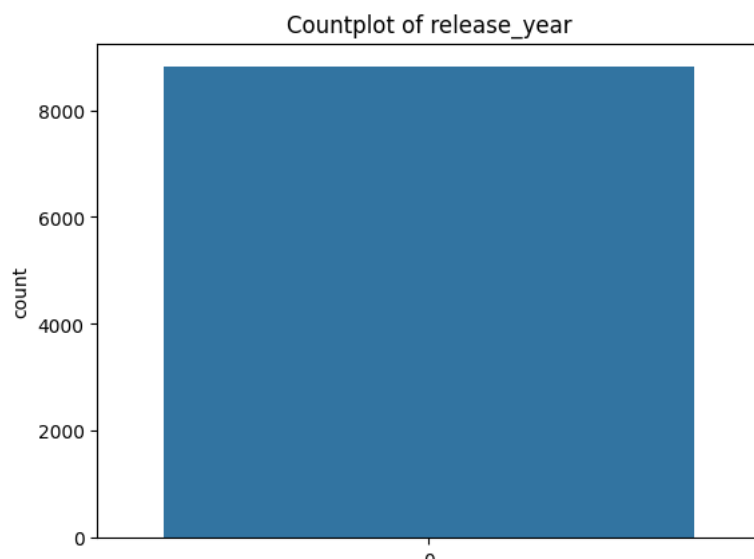
```
show_id      0
type         0
title        0
director    2634
cast         825
country      831
date_added   10
release_year  0
rating       4
duration     3
listed_in    0
description  0
dtype: int64
```

```
continuous_vars = ['duration', 'rating', 'release_year']
```

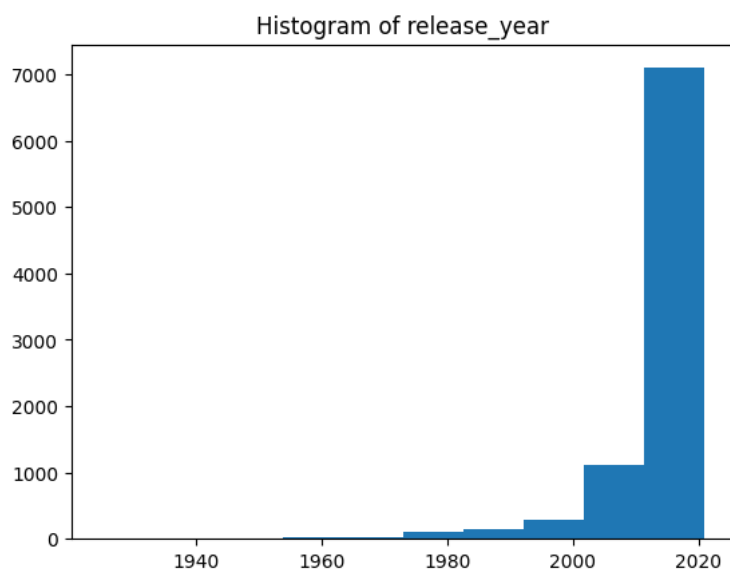
```
for var in continuous_vars:
    # Distplot
    sns.displot(data[var], kde=True)
    plt.title(f"Distribution of {var}")
    plt.show()
```



```
# Countplot
sns.countplot(data[var])
plt.title(f"Countplot of {var}")
plt.show()
```



```
# Histogram
plt.hist(data[var], bins=10)
plt.title(f"Histogram of {var}")
plt.show()
```

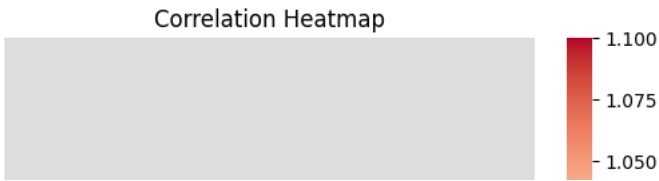


```
correlation_matrix = data.corr()
```

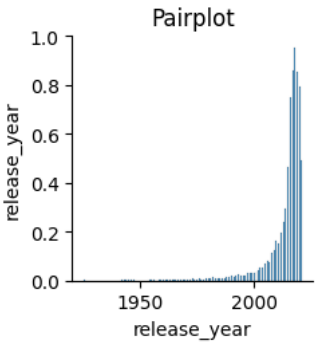
```
<ipython-input-54-19e8f3752a40>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future ve
correlation_matrix = data.corr()
```



```
# Heatmap
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.show()
```

```
# Pairplot
sns.pairplot(data)
plt.title("Pairplot")
plt.show()
```



Summary:

So far we have performed a lot of operations over the dataset to digout some very useful informationfrom it.

- 1. Netflix has more Movies than TV shows.
- 2. Most number of Movies and TV shows are produced by United States, followed by India.
- 3. Most of the content of Netflix is for Mature audience.
- 4. 2018 is the year in which Netflix released a lot of content as compared to other years.
- 5. International Movies and Dramas are the most popular Genres on Netflix followed by Documentaries.
- 6. Visualization showed the exponential increase in the content release over the past years.

Recommendation:

Netflix should creat more content like International Movies and Dramas which are more liked by the audience and should release them on new year. Netflix should also release more content for the mature audience.