In [ ]:
```
import numpy as np
import pandas as pd
```

In [4]:
```
df = pd.read_csv("netflix_titles.csv")
df
```

Out[4]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 23 , 2022	2023	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2023	PG-13	104 min	Comedies, Dramas	A woman adjusting to life after a loss contend...
2	s100	TV Show	On the Verge	NaN	Julie Delpy, Elisabeth Shue, Sarah Jones, Alex...	France, United States	September 7, 2021	2023	TV-MA	1 Season	TV Comedies, TV Dramas	Four women — a chef, a single mom, an heiress ...
3	s1000	Movie	Stowaway	Joe Penna	Anna Kendrick, Toni Collette, Daniel Dae Kim, ...	Germany, United States	April 22, 2021	2023	TV-MA	116 min	Dramas, International Movies, Thrillers	A three-person crew on a mission to Mars faces...
4	s1001	Movie	Wild Dog	Ahishor Solomon	Nagarjuna Akkineni, Dia Mirza, Saiyami Kher, A...	NaN	April 22, 2021	2020	TV-MA	126 min	Action & Adventure, International Movies	A brash but brilliant Indian intelligence agen...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s995	Movie	This Lady Called Life	Kayode Kasum	Bisola Aiyeola, Efa Iwara, Molawa Onajobi, Tin...	Nigeria	April 23, 2021	2020	TV-14	120 min	Dramas, International Movies, Romantic Movies	Abandoned by her family, young single mother A...
8803	s996	Movie	Vizontele	Yılmaz Erdoğan, Ömer Faruk Sorak	Yılmaz Erdoğan, Demet Akbağ, Altan Erkekli, Ce...	Turkey	April 23, 2021	2001	TV-MA	106 min	Comedies, Dramas, International Movies	In 1974, a rural town in Anatolia gets its fir...
8804	s997	Movie	HOMUNCULUS	Takashi Shimizu	Go Ayano, Ryo Narita, Yukino Kishii, Anna Ishi...	Japan	April 22, 2021	2023	TV-MA	116 min	Horror Movies, International Movies, Thrillers	Truth and illusion blurs when a homeless amnes...
8805	s998	TV Show	Life in Color with David Attenborough	NaN	David Attenborough	Australia, United Kingdom	April 22, 2021	2023	TV-PG	1 Season	British TV Shows, Docuseries, International TV...	Using innovative technology, this docuseries e...
8806	s999	Movie	Searching For Sheela	NaN	Ma Anand Sheela	India	April 22, 2021	2023	TV-14	58 min	Documentaries, International Movies	Journalists and fans await Ma Anand Sheela as ...

8807 rows × 12 columns

In [5]:
```
df.shape
```

Out[5]:
```
(8807, 12)
```

In [6]:
```
df.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
```

In [27]:
```
df.describe(include="all")
```

Out[27]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
count	8790	8790	8790	8790	8790	8790	8790	8790.000000	8790	8790	8790	8790
unique	8790	2	8787	4527	7679	749	1765	NaN	14	220	513	8758
top	s1	Movie	15-Aug	unknown director	unknown cast	United States	January 1, 2020	NaN	TV-MA	1 Season	Dramas, International Movies	Paranormal activity at a lush, abandoned prope...
freq	1	6126	2	2621	825	2809	109	NaN	3205	1791	362	4
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2014.669738	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	9.176319	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1925.000000	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2013.000000	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2017.000000	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2020.000000	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2023.000000	NaN	NaN	NaN	NaN

In [7]:
```
df.isnull().sum()
```

Out[7]:
```
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
```

In [8]:
```
df.columns
```

Out[8]:
```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object')
```

In [9]:
```
df.describe()
```

Out[9]:

	release_year
count	8807.000000
mean	2014.665834
std	9.169918
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2020.000000
max	2023.000000

Data Cleaning

In [10]:
```
netflix_movies = df[df['type'] == 'Movie']
movie_added = netflix_movies.groupby('date_added')['show_id'].count()
movie_added
```

Out[10]:
```
date_added
April 1, 2016         3
April 1, 2017        19
April 1, 2018        38
April 1, 2019        21
April 1, 2020        34
                     ..
September 8, 2021     2
September 9, 2016     1
September 9, 2019     1
September 9, 2020     4
September 9, 2021     3
Name: show_id, Length: 1533, dtype: int64
```

Top 10 directors in movies

In [11]:
```
top_10_directors_movie = df['director'][df['type'] == 'Movie'].value_counts().sort_values(ascending=False).iloc[0:10].reset_index()
top_10_directors_movie
```

Out[11]:

	index	director
0	Rajiv Chilaka	19
1	Raúl Campos, Jan Suter	18
2	Suhas Kadav	16
3	Marcus Raboy	15
4	Jay Karas	14
5	Cathy Garcia-Molina	13
6	Jay Chapman	12
7	Youssef Chahine	12
8	Martin Scorsese	12
9	Steven Spielberg	11

top 10 Tv shows

In [12]:
```
top_10_directors_TVShow = df['director'][df['type'] == 'TV Show'].value_counts().sort_values(ascending=False).iloc[0:10].reset_index()
top_10_directors_TVShow
```

Out[12]:

	index	director
0	Alastair Fothergill	3
1	Hsu Fu-chun	2
2	Rob Seidenglanz	2
3	Shin Won-ho	2
4	Ken Burns	2
5	Iginio Straffi	2
6	Stan Lathan	2
7	Alain Brunard	1
8	Jared Hess, Tyler Measom	1
9	Maribel Sánchez-Maroto	1

In [13]:
```
top_10_countries_TVShow = df['country'][df['type'] == 'TV Show'].value_counts().sort_values(ascending=False).iloc[0:10].reset_index()
top_10_countries_TVShow
```

Out[13]:

	index	country
0	United States	760
1	United Kingdom	213
2	Japan	169
3	South Korea	158
4	India	79
5	Taiwan	68
6	Canada	59
7	France	49
8	Spain	48
9	Australia	48

In [14]:
```
top_10_countries_movie = df['country'][df['type'] == 'Movie'].value_counts().sort_values(ascending=False).iloc[0:10].reset_index()
top_10_countries_movie
```

Out[14]:

	index	country
0	United States	2058
1	India	893
2	United Kingdom	206
3	Canada	122
4	Spain	97
5	Egypt	92
6	Nigeria	86
7	Indonesia	77
8	Turkey	76
9	Japan	76

In [15]:
```
top_10_rating_movie = df['rating'][df['type'] == 'Movie'].value_counts().sort_values(ascending=False).iloc[0:10].reset_index()
top_10_rating_movie
```

Out[15]:

	index	rating
0	TV-MA	2062
1	TV-14	1427
2	R	797
3	TV-PG	540
4	PG-13	490
5	PG	287
6	TV-Y7	139
7	TV-Y	131
8	TV-G	126
9	NR	75

In [16]:
```
top_10_rating_tvshow = df['rating'][df['type'] == 'TV Show'].value_counts().sort_values(ascending=False).iloc[0:10].reset_index()
top_10_rating_tvshow
```

Out[16]:

	index	rating
0	TV-MA	1145
1	TV-14	733
2	TV-PG	323
3	TV-Y7	195
4	TV-Y	176
5	TV-G	94
6	NR	5
7	R	2
8	TV-Y7-FV	1

In [17]:
```
dff = df[['type','release_year']]
dff = dff.rename(columns = {'release_year' : 'Release Year'})
dff2 = dff.groupby(['Release Year','type']).size().reset_index(name='Total Content')
dff2 = dff2[dff2['Release Year']>=2011]
dff2
```

Out[17]:

	Release Year	type	Total Content
97	2011	Movie	145
98	2011	TV Show	40
99	2012	Movie	173
100	2012	TV Show	64
101	2013	Movie	225
102	2013	TV Show	63
103	2014	Movie	264
104	2014	TV Show	88
105	2015	Movie	398
106	2015	TV Show	162
107	2016	Movie	658
108	2016	TV Show	244
109	2017	Movie	767
110	2017	TV Show	265
111	2018	Movie	767
112	2018	TV Show	380
113	2020	Movie	516
114	2020	TV Show	436
115	2022	Movie	633
116	2022	TV Show	397
117	2023	Movie	278
118	2023	TV Show	315

In [18]:
```
df.isnull().sum()
```

Out[18]:
```
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
```

In [19]:
```
df['director']= df['director'].fillna('unknown director')
df['country']= df['country'].fillna('unknown country')
df.head()
```

Out[19]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 23 , 2022	2023	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2023	PG-13	104 min	Comedies, Dramas	A woman adjusting to life after a loss contend...
2	s100	TV Show	On the Verge	unknown director	Julie Delpy, Elisabeth Shue, Sarah Jones, Alex...	France, United States	September 7, 2021	2023	TV-MA	1 Season	TV Comedies, TV Dramas	Four women — a chef, a single mom, an heiress ...
3	s1000	Movie	Stowaway	Joe Penna	Anna Kendrick, Toni Collette, Daniel Dae Kim, ...	Germany, United States	April 22, 2021	2023	TV-MA	116 min	Dramas, International Movies, Thrillers	A three-person crew on a mission to Mars faces...
4	s1001	Movie	Wild Dog	Ahishor Solomon	Nagarjuna Akkineni, Dia Mirza, Saiyami Kher, A...	unknown country	April 22, 2021	2020	TV-MA	126 min	Action & Adventure, International Movies	A brash but brilliant Indian intelligence agen...

In [20]:
```
df.isnull().sum()
```

Out[20]:
```
show_id           0
type              0
title             0
director          0
cast            825
country           0
date_added       10
release_year      0
rating            4
duration          3
listed_in         0
description       0
dtype: int64
```

In [30]:
```
# Creating a function to categorise records by audience type adult or children based on 'rating'
def categaudience (rating):
    adult_rating = ['TV-MA', 'R', 'NC-17', 'PG-13', 'TV-14']
    children_rating = ['PG', 'TV-PG', 'TV-Y', 'TV-Y7', 'TV-G', 'G', 'TV-Y7-FV', 'Not Rated']
    if rating in adult_rating:
        return 'Adult'
    if rating in children_rating:
        return 'Children'
    else:
        return 'Not Rated'

# Applying categorisation function to create the 'audience' column
df['audience'] = df['rating'].apply(categaudience)
```

In [31]:
```
df.isnull().sum()
```

Out[31]:
```
show_id         0
type            0
title           0
director        0
cast            0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
description     0
audience        0
dtype: int64
```

In [32]:
```
df.dropna(inplace=True,axis=0,subset=['duration'])
```

In [33]:
```
df.isnull().sum()
```

Out[33]:
```
show_id         0
type            0
title           0
director        0
cast            0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
description     0
audience        0
dtype: int64
```

In [35]:
```
df.dropna(inplace=True, axis=0, subset=['date_added','rating'])
```

In [36]:
```
df.isnull().sum()
```

Out[36]:
```
show_id         0
type            0
title           0
director        0
cast            0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
description     0
audience        0
dtype: int64
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: