In [8]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [10]:
```
df = pd.read_csv('netflix1.csv')
```

In [12]:
```
df.head()
```

Out[12]:

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies

In [14]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8790 entries, 0 to 8789
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8790 non-null   object
 1   type          8790 non-null   object
 2   title         8790 non-null   object
 3   director      8790 non-null   object
 4   country       8790 non-null   object
 5   date_added    8790 non-null   object
 6   release_year  8790 non-null   int64 
 7   rating        8790 non-null   object
 8   duration      8790 non-null   object
 9   listed_in     8790 non-null   object
dtypes: int64(1), object(9)
memory usage: 686.8+ KB
```

In [16]:
```
df['listed_in'].head()
```

Out[16]:
```
0                                  Documentaries
1    Crime TV Shows, International TV Shows, TV Act...
2                  TV Dramas, TV Horror, TV Mysteries
3                  Children & Family Movies, Comedies
4     Dramas, Independent Movies, International Movies
Name: listed_in, dtype: object
```

In [18]:
```
df['type'].head()
```

Out[18]:
```
0      Movie
1    TV Show
2    TV Show
3      Movie
4      Movie
Name: type, dtype: object
```

In [20]:
```
df['listed_in']
```

Out[20]:
```
0                                       Documentaries
1       Crime TV Shows, International TV Shows, TV Act...
2                     TV Dramas, TV Horror, TV Mysteries
3                     Children & Family Movies, Comedies
4        Dramas, Independent Movies, International Movies
                              ...
8785                   International TV Shows, TV Dramas
8786                                        Kids' TV
8787    International TV Shows, Romantic TV Shows, TV ...
8788                                        Kids' TV
8789                                        Kids' TV
Name: listed_in, Length: 8790, dtype: object
```

In [22]:
```
df.duplicated().sum()
```

Out[22]:
```
0
```

In [24]:
```
df.duplicated()
```

Out[24]:
```
0       False
1       False
2       False
3       False
4       False
        ...
8785    False
8786    False
8787    False
8788    False
8789    False
Length: 8790, dtype: bool
```

In [26]:
```
df.describe()
```

Out[26]:

	release_year
count	8790.000000
mean	2014.183163
std	8.825466
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

Data Cleaning

In [12]:
```
print(df.isnull().sum())
```

```
show_id         0
type            0
title           0
director        0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
dtype: int64
```

In [22]:
```
print(df.drop_duplicates(inplace=True))
```

```
None
```

In [24]:
```
print(df.dtypes)
```

```
show_id         object
type            object
title           object
director        object
country         object
date_added      object
release_year     int64
rating          object
duration        object
listed_in       object
dtype: object
```

In [34]:
```
df['date_added']=pd.to_datetime(df['date_added'])
```

In [36]:
```
print(df.dtypes)
```

```
show_id                 object
type                    object
title                   object
director                object
country                 object
date_added      datetime64[ns]
release_year             int64
rating                  object
duration                object
listed_in               object
dtype: object
```

Exploratory Data Analysis

In [ ]:
```
# Count the number of movies and TV shows
```

In [40]:
```
print(df['type'].value_counts())
```

```
type
Movie      6126
TV Show    2664
Name: count, dtype: int64
```