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

```
warnings.filterwarnings("ignore")


%matplotlib inline
```

```
df=pd.read_csv('https://raw.githubusercontent.com/allenkong221/netflix-titles-dataset/refs/hea
df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	du
0	81145628	Movie	Norm of the North: King Sized Adventure	Richard Finn, Tim Maltby	Alan Marriott, Andrew Toth, Brian Dobson, Cole...	United States, India, South Korea, China	September 9, 2019	2019	TV-PG	
1	80117401	Movie	Jandino: Whatever it Takes	NaN	Jandino Asporaat	United Kingdom	September 9, 2016	2016	TV-MA	
2	70234439	TV Show	Transformers Prime	NaN	Peter Cullen, Sumalee Montano, Frank Welker, J...	United States	September 8, 2018	2013	TV-Y7-FV	1 S
3	80058654	TV Show	Transformers: Robots in Disguise	NaN	Will Friedle, Darren Criss, Constance Zimmer, ...	United States	September 8, 2018	2016	TV-Y7	1 S
4	80125979	Movie	#realityhigh	Fernando Lebrija	Nesta Cooper, Kate Walsh, John Michael Higgins...	United States	September 8, 2017	2017	TV-14	

```
df.shape
```

```
(6234, 12)
```

```
df.info
```

```
<bound method DataFrame.info of        show_id     type
title  \
0     81145628    Movie      Norm of the North: King Sized Adventure
1     80117401    Movie                    Jandino: Whatever it Takes
2     70234439  TV Show                           Transformers Prime
3     80058654  TV Show             Transformers: Robots in Disguise
4     80125979    Movie                                  #realityhigh
...        ...      ...                                           ...
6229  80000063  TV Show                                Red vs. Blue
6230  70286564  TV Show                                       Maron
6231  80116008    Movie       Little Baby Bum: Nursery Rhyme Friends
6232  70281022  TV Show  A Young Doctor's Notebook and Other Stories
6233  70153404  TV Show                                     Friends

                     director  \
0     Richard Finn, Tim Maltby
1                          NaN
2                          NaN
3                          NaN
4              Fernando Lebrija
...                         ...
6229                       NaN
6230                       NaN
6231                       NaN
6232                       NaN
6233                       NaN

                                                cast  \
0     Alan Marriott, Andrew Toth, Brian Dobson, Cole...
1                                     Jandino Asporaat
2     Peter Cullen, Sumalee Montano, Frank Welker, J...
3     Will Friedle, Darren Criss, Constance Zimmer, ...
4     Nesta Cooper, Kate Walsh, John Michael Higgins...
...                                                 ...
6229  Burnie Burns, Jason Saldaña, Gustavo Sorola, G...
6230  Marc Maron, Judd Hirsch, Josh Brener, Nora Zeh...
6231                                                NaN
6232  Daniel Radcliffe, Jon Hamm, Adam Godley, Chris...
6233  Jennifer Aniston, Courteney Cox, Lisa Kudrow, ...

                                     country         date_added  \
0     United States, India, South Korea, China  September 9, 2019
1                                United Kingdom  September 9, 2016
2                                 United States  September 8, 2018
3                                 United States  September 8, 2018
4                                 United States  September 8, 2017
...                                         ...               ...
6229                              United States               NaN
6230                              United States               NaN
6231                                        NaN               NaN
6232                             United Kingdom               NaN
6233                              United States               NaN

      release_year    rating    duration  \
0             2019     TV-PG      90 min
1             2016     TV-MA      94 min
2             2013  TV-Y7-FV    1 Season
```

```
3             2016     TV-Y7    1 Season
4             2017     TV-14      99 min
...            ...       ...         ...
6229          2015        NR  13 Seasons
6230          2016     TV-MA   4 Seasons
6231          2016       NaN      60 min
6232          2013     TV-MA   2 Seasons
6233          2003     TV-14  10 Seasons

                                          listed_in  \
0                   Children & Family Movies, Comedies
1                                  Stand-Up Comedy
2                                         Kids' TV
3                                         Kids' TV
4                                         Comedies
...                                            ...
6229  TV Action & Adventure, TV Comedies, TV Sci-Fi ...
6230                                    TV Comedies
6231                                         Movies
6232           British TV Shows, TV Comedies, TV Dramas
6233                     Classic & Cult TV, TV Comedies

                                          description
0     Before planning an awesome wedding for his gra...
1     Jandino Asporaat riffs on the challenges of ra...
2     With the help of three human allies, the Autob...
3     When a prison ship crash unleashes hundreds of...
4     When nerdy high schooler Dani finally attracts...
...                                                 ...
6229  This parody of first-person shooter games, mil...
6230  Marc Maron stars as Marc Maron, who interviews...
6231  Nursery rhymes and original music for children...
6232  Set during the Russian Revolution, this comic ...
6233  This hit sitcom follows the merry misadventure...

[6234 rows x 12 columns]>
```

```
df.describe()
```

	show_id	release_year
count	6.234000e+03	6234.00000
mean	7.670368e+07	2013.35932
std	1.094296e+07	8.81162
min	2.477470e+05	1925.00000
25%	8.003580e+07	2013.00000
50%	8.016337e+07	2016.00000
75%	8.024489e+07	2018.00000
max	8.123573e+07	2020.00000

```
## missing values
df.isnull().sum()
```

```
show_id            0
type               0
title              0
director        1969
cast             570
country          476
date_added        11
release_year       0
rating            10
duration           0
listed_in          0
description        0
dtype: int64
```

```
df.duplicated('title')
```

```
0       False
1       False
2       False
3       False
4       False
        ...
6229    False
6230    False
6231     True
6232    False
6233    False
Length: 6234, dtype: bool
```

```
df['country'].fillna('Unknown' , inplace=True)


df.info()
df.shape
```

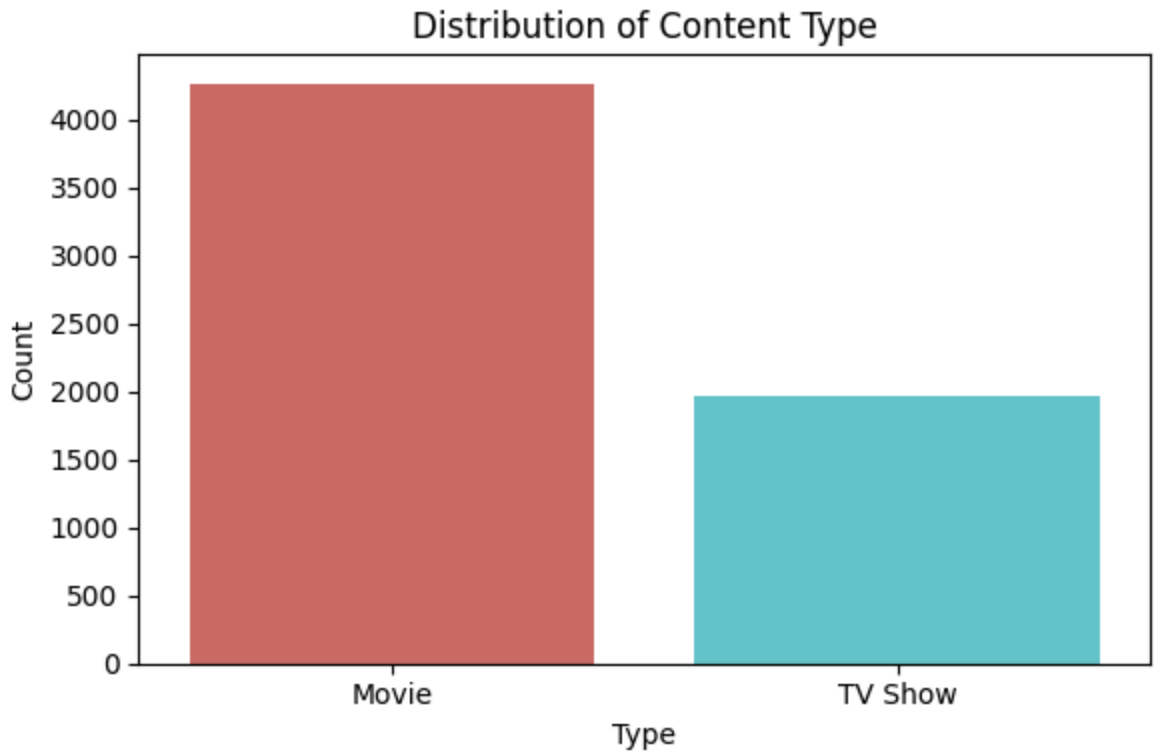
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6234 entries, 0 to 6233
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       6234 non-null   int64 
 1   type          6234 non-null   object
 2   title         6234 non-null   object
 3   country       6234 non-null   object
 4   date_added    6223 non-null   object
 5   release_year  6234 non-null   int64 
 6   rating        6224 non-null   object
 7   duration      6234 non-null   object
 8   listed_in     6234 non-null   object
 9   description   6234 non-null   object
dtypes: int64(2), object(8)
memory usage: 487.2+ KB
(6234, 10)
```

```
## EDA

1.what is the distribution of content types (movies vs tvshows) ?
2. what are the most common genres?
3. how has netflix content changed over time?
4. which countries produce the most content?
```

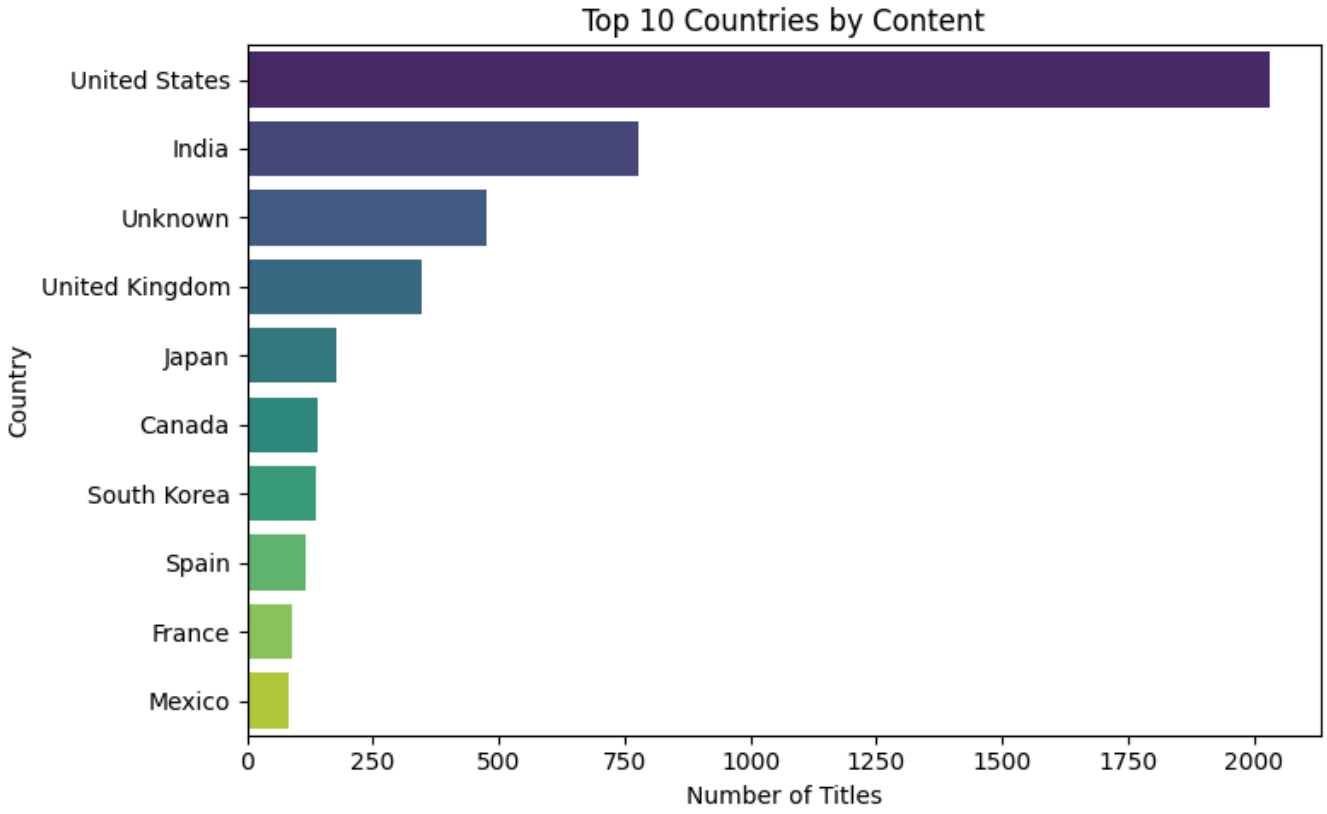
```
df['date_added']= pd.to_datetime(df['date_added'], errors='coerce')
```

```
df['year_added'] = df['date_added'].dt.year
df['month_added'] = df['date_added'].dt.month
```

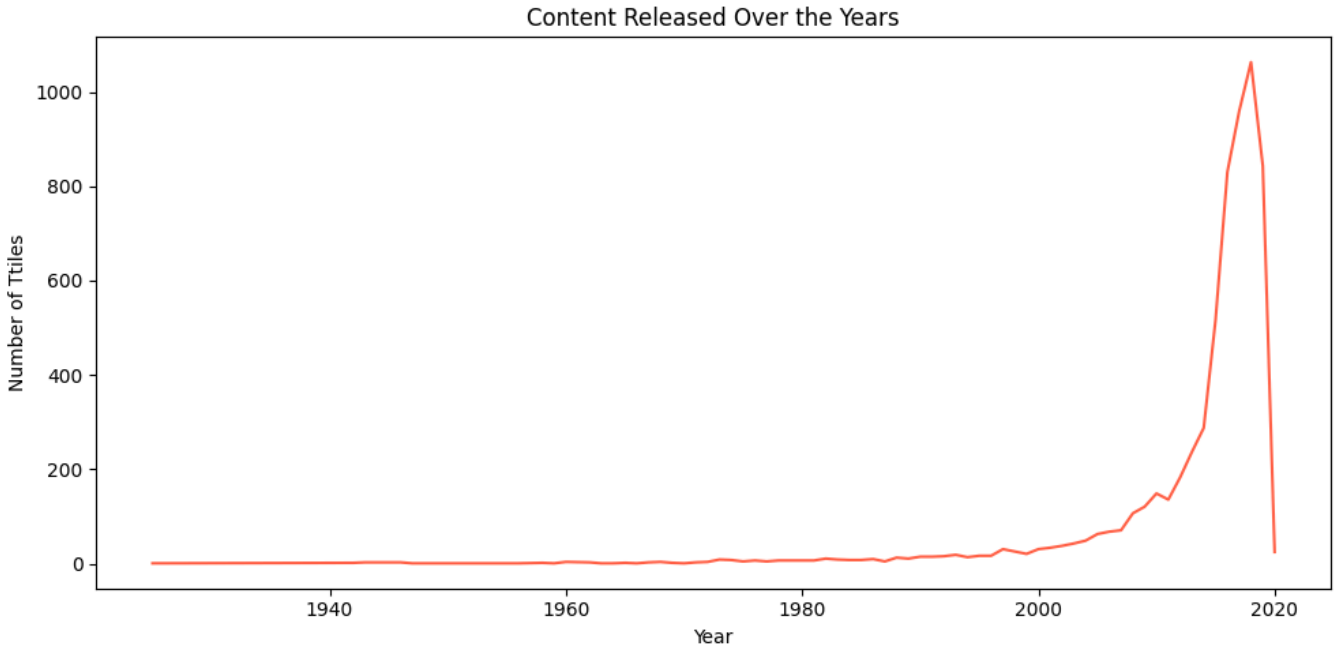
```
##1.content type count

plt.figure(figsize=(6,4))
sns.countplot(x='type', data=df, palette='hls')
plt.title('Distribution of Content Type')
plt.xlabel('Type')
plt.ylabel('Count')
plt.tight_layout()
plt.show()
```



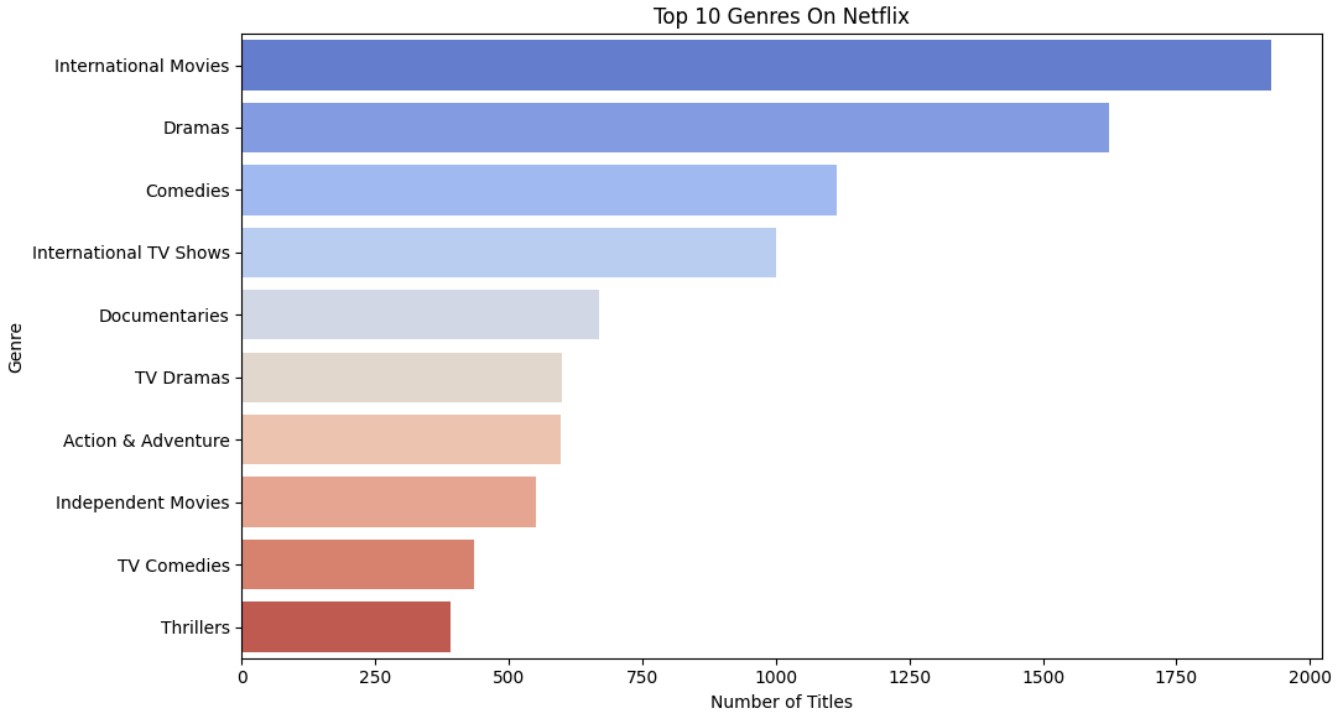

```
## 2. top 10 countries with most content

top_countries = df['country'].value_counts().head(10)
plt.figure(figsize=(8,5))
sns.barplot(x=top_countries.values , y=top_countries.index, palette='viridis')
plt.title('Top 10 Countries by Content')
plt.xlabel('Number of Titles')
plt.ylabel('Country')
plt.tight_layout()
plt.show()
```



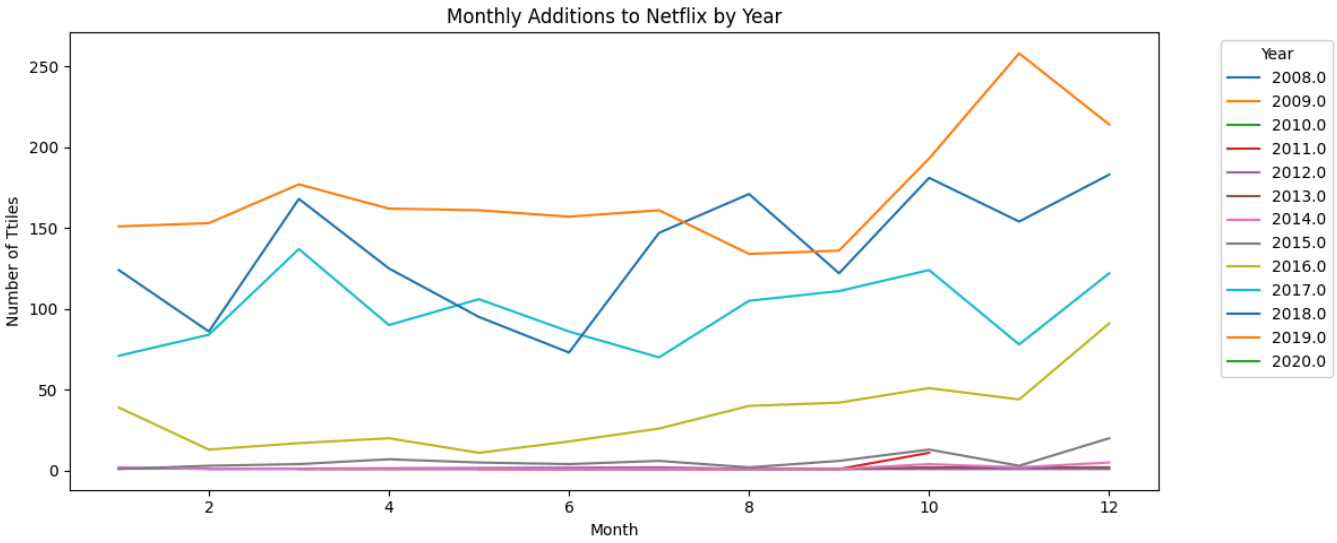

```
## 3. Content over years

plt.figure(figsize=(10,5))
df['release_year'].value_counts().sort_index().plot(kind='line' , color='tomato')
plt.title('Content Released Over the Years')
plt.xlabel('Year')
plt.ylabel('Number of Ttiles')
plt.tight_layout()
plt.show()
```



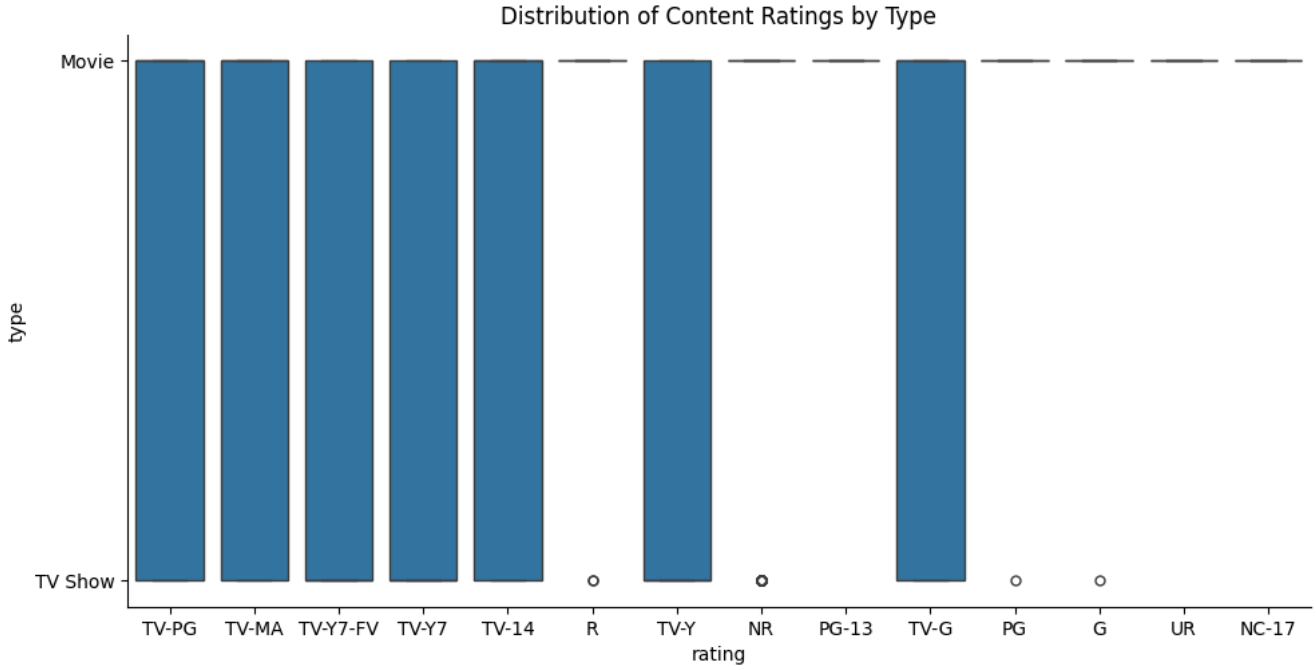
```
## 4. Most Common Genres

genres = df['listed_in'].str.split(', ').explode()
top_genres = genres.value_counts().head(10)
plt.figure(figsize=(11,6))
sns.barplot(x=top_genres.values , y=top_genres.index, palette='coolwarm')
plt.title('Top 10 Genres On Netflix')
plt.xlabel('Number of Titles')
plt.ylabel('Genre')
plt.tight_layout()
plt.show()
```




```
## 5. Monthly additions to Netflix

df['month_added'] = df['date_added'].dt.month
df['year_added'] = df['date_added'].dt.year
monthly_counts = df.groupby(['year_added', 'month_added']).size().reset_index(name='count')

plt.figure(figsize=(12,5))
sns.lineplot(data=monthly_counts, x='month_added', y='count', hue='year_added' , palette='tab1
plt.title('Monthly Additions to Netflix by Year')
plt.xlabel('Month')
plt.ylabel('Number of Ttiles')
plt.legend(title='Year' , bbox_to_anchor=(1.05, 1), loc ='upper left')
plt.tight_layout()
plt.show()
```




```
sns.catplot(x='rating' , y='type' , data=df , kind='box' , height=5, aspect=2)
plt.title('Distribution of Content Ratings by Type')
plt.show()
```



FEATURE ENGINEERING

```
# 1. content Age

df['content_age'] = 2025 - df['release_year']
```

```
# 2. year and month added
df['year_added'] = df['date_added'].dt.year
df['month_added'] = df['date_added'].dt.month
```

```
# 3. title length

df['title_length'] = df['title'].apply(len)
```

```
# 4. main genre (first listed)

df['main genre']= df['listed_in'].str.split(', ').str[0]
```

```
# 5. is international

df['is_international']= df['country'].apply(lambda x: 'United States' not in x)
```

```
# 6. is movie

df['is_movie'] = df['type'].apply(lambda x: 1 if x == 'Movie' else 0)
```

```
# to check if cast column was dropped
df.columns
```

```
Index(['show_id', 'type', 'title', 'country', 'date_added', 'release_year',
       'rating', 'duration', 'listed_in', 'description', 'year_added',
       'month_added', 'content_age', 'title_length', 'main genre',
       'is_international', 'is_movie'],
      dtype='object')
```

```
df.columns = df.columns.str.lower().str.replace(" ", "_")
```

```
print(df[['title', 'type', 'release_year', 'content_age', 'year_added', 'is_movie']].head())
```

```
                                  title     type  release_year  \
0  Norm of the North: King Sized Adventure    Movie          2019
1               Jandino: Whatever it Takes    Movie          2016
2                     Transformers Prime  TV Show          2013
3         Transformers: Robots in Disguise  TV Show          2016
4                              #realityhigh    Movie          2017

   content_age  year_added  is_movie
0            6      2019.0         1
1            9      2016.0         1
2           12      2018.0         0
3            9      2018.0         0
4            8      2017.0         1
```

```
df.head()
```

	show_id	type	title	country	date_added	release_year	rating	duration	listed_in	des
0	81145628	Movie	Norm of the North: King Sized Adventure	United States, India, South Korea, China	2019-09-09	2019	TV-PG	90 min	Children & Family Movies, Comedies	pla a for
1	80117401	Movie	Jandino: Whatever it Takes	United Kingdom	2016-09-09	2016	TV-MA	94 min	Stand-Up Comedy	rif ch
2	70234439	TV Show	Transformers Prime	United States	2018-09-08	2013	TV-Y7-FV	1 Season	Kids' TV	a
3	80058654	TV Show	Transformers: Robots in Disguise	United States	2018-09-08	2016	TV-Y7	1 Season	Kids' TV	pr u h
4	80125979	Movie	#realityhigh	United States	2017-09-08	2017	TV-14	99 min	Comedies	ne Da