

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

```
data=pd.read_csv('netflix1.csv')
```

data

	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
...
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	TV-PG	2 Seasons	International TV Shows, TV Dramas

Next steps: [Generate code with data](#) [New interactive sheet](#)

data.head()

	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

Next steps: [Generate code with data](#) [New interactive sheet](#)

data.tail()

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	TV-PG	2 Seasons	International TV Shows, TV Dramas
8786	s8798	TV Show	Zak Storm	Not Given	United States	9/13/2018	2016	TV-Y7	3 Seasons	Kids' TV
8787	s8801	TV Show	Zindagi Gulzar Hai	Not Given	Pakistan	12/15/2016	2012	TV-PG	1 Season	International TV Shows, Romantic TV Shows, TV ...

data.sample()

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
1184	s1001	Movie	Dolly Parton's ...	Debbie	United	11/22/2020	2020	TV-PG	90 min	Children & Family ...

data.shape

(8790, 10)

data.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
```

```
data.describe
```

```
pandas.core.generic.NDFrame.describe
def describe(percentiles=None, include=None, exclude=None) -> Self
```

</usr/local/lib/python3.12/dist-packages/pandas/core/generic.py>
Generate descriptive statistics.

Descriptive statistics include those that summarize the central tendency, dispersion and shape of a dataset's distribution, excluding ``NaN`` values.

```
data.columns
```

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

```
data.index
```

```
RangeIndex(start=0, stop=8790, step=1)
```

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

```

      0
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
```

```
data.rating.unique()
```

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

```
data.duplicated().sum()
```

```
np.int64(0)
```

```
data['show_id']=data['show_id'].str.replace('s', '')
data.head()
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in	
0	1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries	
1	3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	
2	6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	

Next steps: [Generate code with data](#) [New interactive sheet](#)

```
data.tail()
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in	
8785	8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	TV-PG	2 Seasons	International TV Shows, TV Dramas	
8786	8798	TV Show	Zak Storm	Not Given	United States	9/13/2018	2016	TV-Y7	3 Seasons	Kids' TV	
8787	8801	TV Show	Zindagi Gulzar Hai	Not Given	Pakistan	12/15/2016	2012	TV-PG	1 Season	International TV Shows, Romantic TV Shows, TV ...	

```
data.groupby('country')['country'].count().sort_values(ascending=False)
```

country	
country	
United States	3240
India	1057
United Kingdom	638
Pakistan	421
Not Given	287
...	...
Slovenia	1
Puerto Rico	1
Somalia	1
West Germany	1
Zimbabwe	1

86 rows × 1 columns

dtype: int64

```
data.groupby('country')['country'].count().sort_values(ascending=False).head()
```

country	
country	
United States	3240
India	1057
United Kingdom	638
Pakistan	421
Not Given	287

dtype: int64

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

```

count
type
Movie    6126
TV Show  2664

dtype: int64

```

```
data.pivot_table(index='type', values='release_year', aggfunc='count')
```

```

release_year
type
Movie          6126
TV Show        2664

```

```
data.loc[data['release_year']==2020]
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries
25	17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	Pedro de Echave Garcia, Pablo Azorin Williams	Not Given	9/22/2021	2020	TV-MA	67 min	Documentaries, International Movies
36	35	TV Show	Tayo and Little Wizards	Not Given	Pakistan	9/17/2021	2020	TV-Y7	1 Season	Kids' TV
47	190	TV Show	Bread Barbershop	Not Given	Pakistan	8/28/2021	2020	TV-Y	2 Seasons	Kids' TV, TV Comedies
58	48	TV Show	The Smart Money Woman	Bunmi Ajakaiye	South Africa	9/16/2021	2020	TV-MA	1 Season	International TV Shows, Romantic TV Shows, TV ...
...
7484	3288	TV Show	Maradona in Mexico	Not Given	Argentina	11/13/2019	2020	TV-MA	1 Season	Docuseries, Spanish-Language TV Shows
7517	3370	TV Show	BoJack Horseman	Not Given	United States	10/25/2019	2020	TV-MA	6 Seasons	TV Comedies
7537	3434	TV Show	The Hook Up Plan	Not Given	France	10/11/2019	2020	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...
8005	3400	TV	United	...	2020	TV-Y	3	Kids' TV. Korean

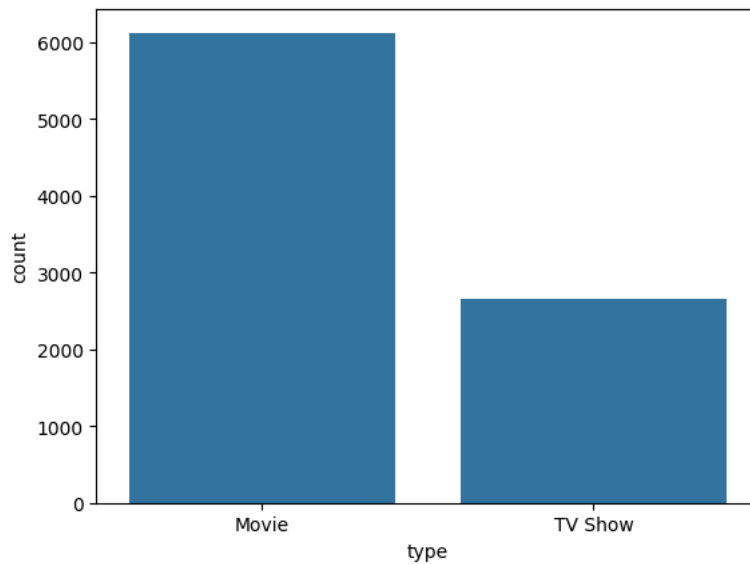
```
data.iloc[100:110]
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
100	370	TV Show	Myth & Mogul: John DeLorean	Not Given	Pakistan	7/30/2021	2021	TV-14	1 Season	British TV Shows, Crime TV Shows, Docuseries
101	377	TV Show	Transformers: War for Cybertron: Kingdom	Not Given	Pakistan	7/29/2021	2021	TV-Y7	1 Season	Anime Series
102	380	TV Show	Tattoo Redo	Not Given	Pakistan	7/28/2021	2021	TV-MA	1 Season	Reality TV
103	382	TV Show	The Snitch Cartel: Origins	Not Given	Pakistan	7/28/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, Spanis...
104	398	TV Show	Feels Like Ishq	Not Given	Pakistan	7/23/2021	2021	TV-MA	1 Season	International TV Shows, Romantic TV Shows, TV ...

```

sns.countplot(x='type', data=data)
plt.show()

```



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

count	
country	
United States	3240
India	1057
United Kingdom	638
Pakistan	421
Not Given	287
...	...
Luxembourg	1
Senegal	1
Belarus	1
Puerto Rico	1
Cyprus	1

86 rows × 1 columns

dtype: int64

```
top_10=data['country'].value_counts().nlargest(10)  
top_10
```

count	
country	
United States	3240
India	1057
United Kingdom	638
Pakistan	421
Not Given	287
Canada	271
Japan	259
South Korea	214
France	213
Spain	182

dtype: int64

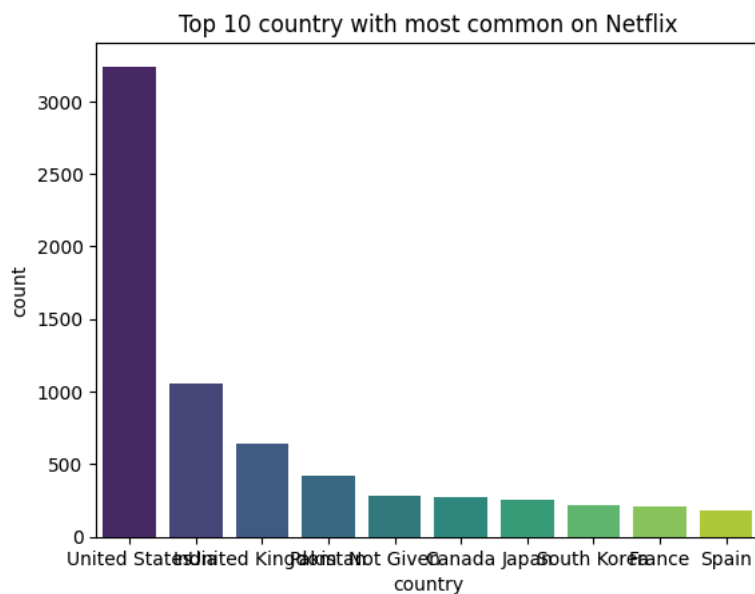
```
sns.barplot(x=top_10.index,y=top_10.values,palette='viridis')  
plt.xlabel('country')  
plt.ylabel('count')
```

```
plt.title('Top 10 country with most common on Netflix')
plt.show()
```

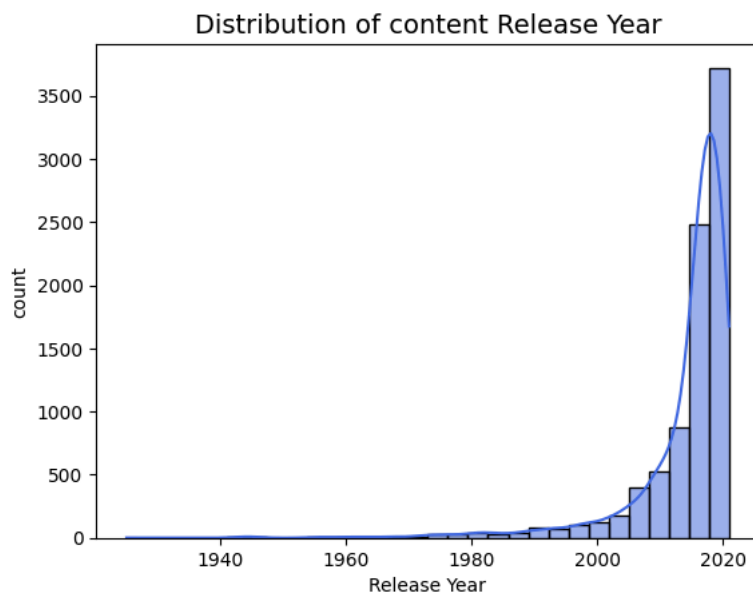
/tmp/ipython-input-169615352.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set

```
sns.barplot(x=top_10.index,y=top_10.values,palette='viridis')
```



```
sns.histplot(data['release_year'],bins=30, kde=True, color='royalblue')
plt.title('Distribution of content Release Year', fontsize=14)
plt.xlabel('Release Year')
plt.ylabel('count')
plt.show()
```

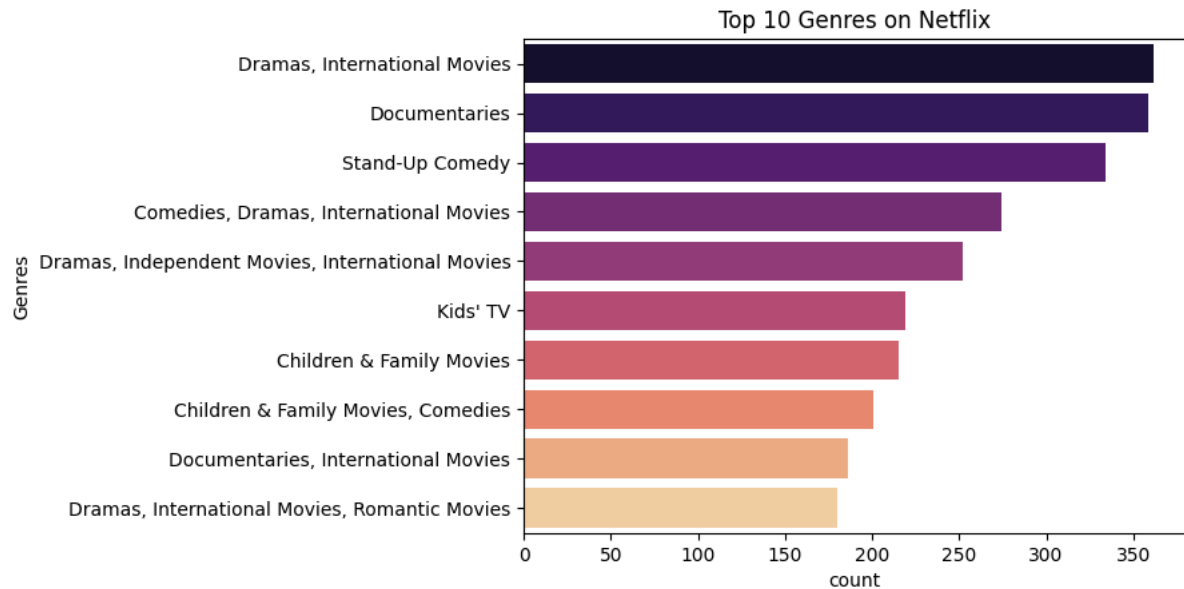


```
top_genres = data['listed_in'].value_counts().head(10)
sns.barplot(x=top_genres.index, y=top_genres.values, palette='magma')
plt.xlabel('count')
plt.ylabel('Genres')
plt.title('Top 10 Genres on Netflix')
plt.show()
```

```
/tmp/ipython-input-2342902823.py:2: FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set

```
sns.barplot(x=top_genres.values, y=top_genres.index, palette='magma')
```

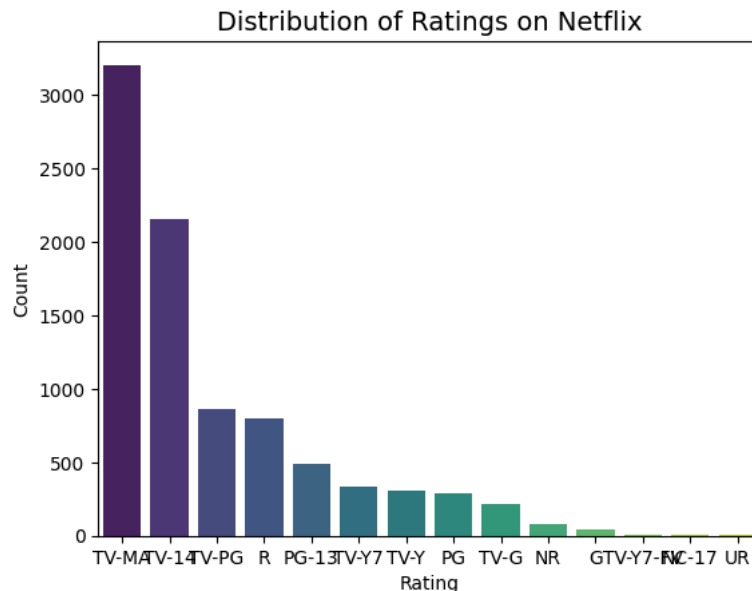


```
rating_counts = data['rating'].value_counts()
sns.barplot(x=rating_counts.index, y=rating_counts.values, palette='viridis')
plt.xlabel('Rating')
plt.ylabel('Count')
plt.title('Distribution of Ratings on Netflix', fontsize=14)
plt.show()
```

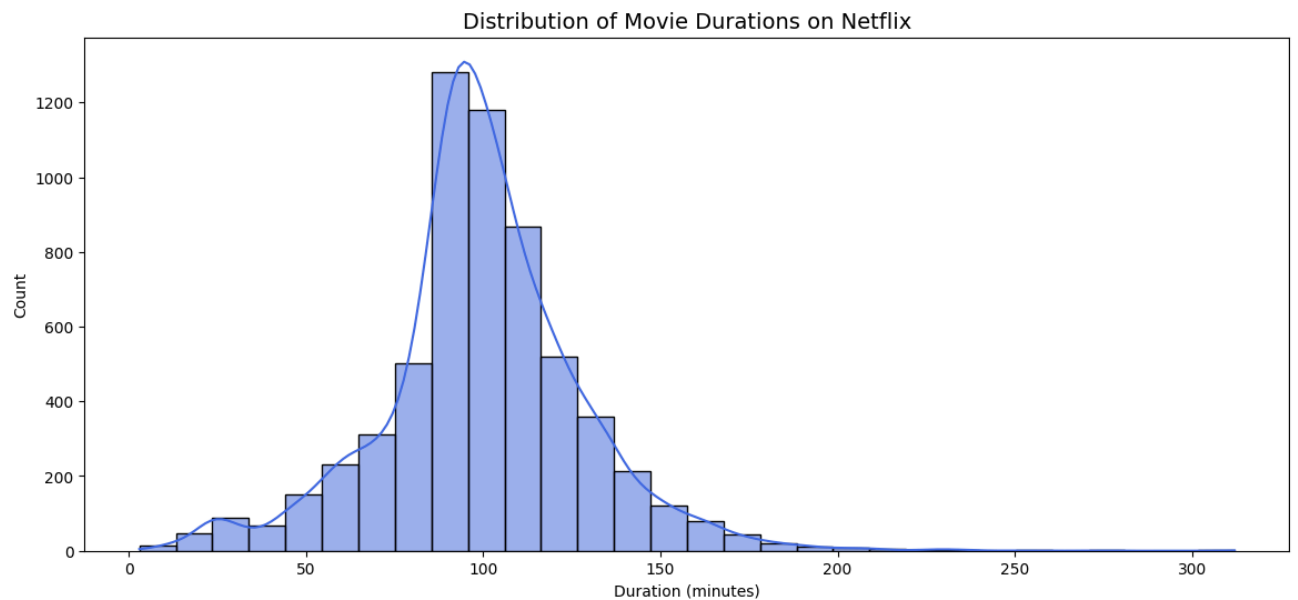
```
/tmp/ipython-input-252468191.py:2: FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set

```
sns.barplot(x=rating_counts.index, y=rating_counts.values, palette='viridis')
```



```
movies_df = data[data['type'] == 'Movie'].copy()
movies_df['duration'] = movies_df['duration'].str.replace('min', '').astype(float)
plt.figure(figsize=(14, 6))
sns.histplot(movies_df['duration'], bins=30, kde=True, color='royalblue')
plt.title('Distribution of Movie Durations on Netflix', fontsize=14)
plt.xlabel('Duration (minutes)')
plt.ylabel('Count')
plt.show()
```

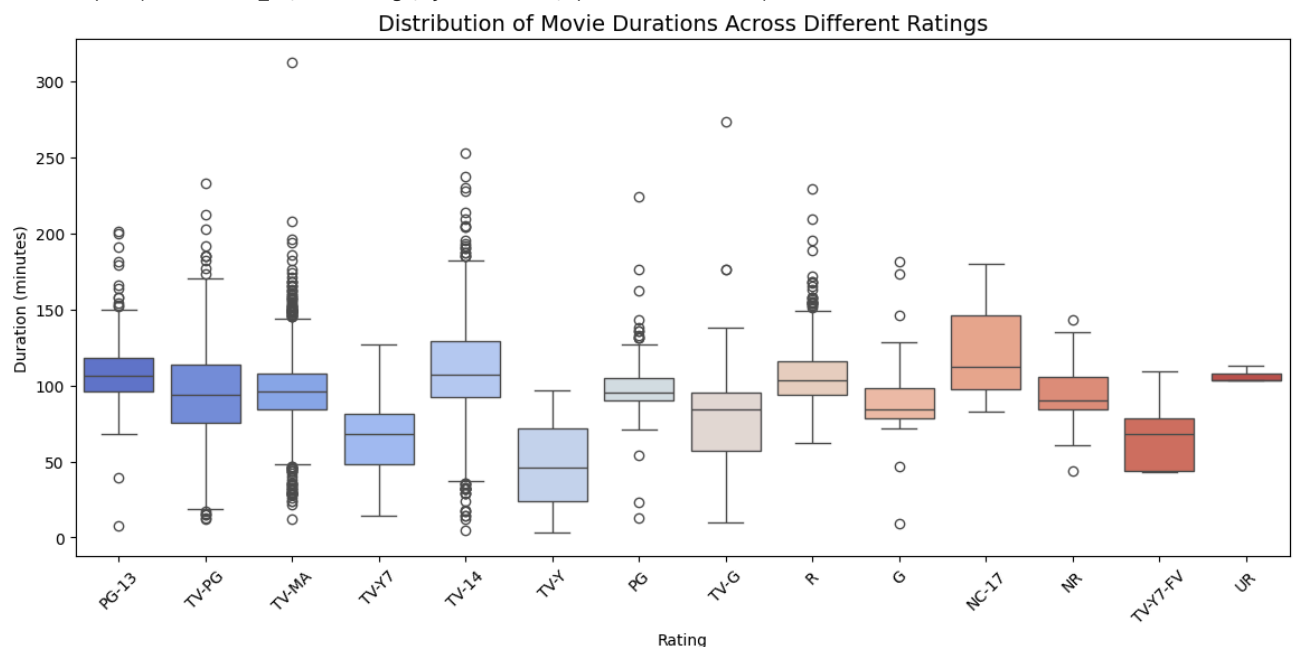


```
plt.figure(figsize=(14,6))
sns.boxplot(data=movies_df, x='rating', y='duration', palette='coolwarm')
plt.title('Distribution of Movie Durations Across Different Ratings', fontsize=14)
plt.xlabel('Rating')
plt.ylabel('Duration (minutes)')
plt.xticks(rotation=45)
plt.show()
```

/tmp/ipython-input-596010664.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set the `palette` parameter.

```
sns.boxplot(data=movies_df, x='rating', y='duration', palette='coolwarm')
```



```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
a=['show_id', 'type', 'director', 'title', 'country', 'date_added', 'rating', 'listed_in']
for i in a:
    data[i]=le.fit_transform(data[i])
```



```
data.head()
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	0	0	1972	2294	80	1656	2020	4	90 min	273
1	2222	1	2644	2104	20	1651	2021	8	1 Season	241
2	5551	1	4558	2865	80	1651	2021	8	1 Season	498
3	445	0	1657	627	6	1642	2021	9	91 min	124
4	7761	0	6043	1503	80	1651	1993	8	125 min	318

Next steps: [Generate code with data](#) [New interactive sheet](#)

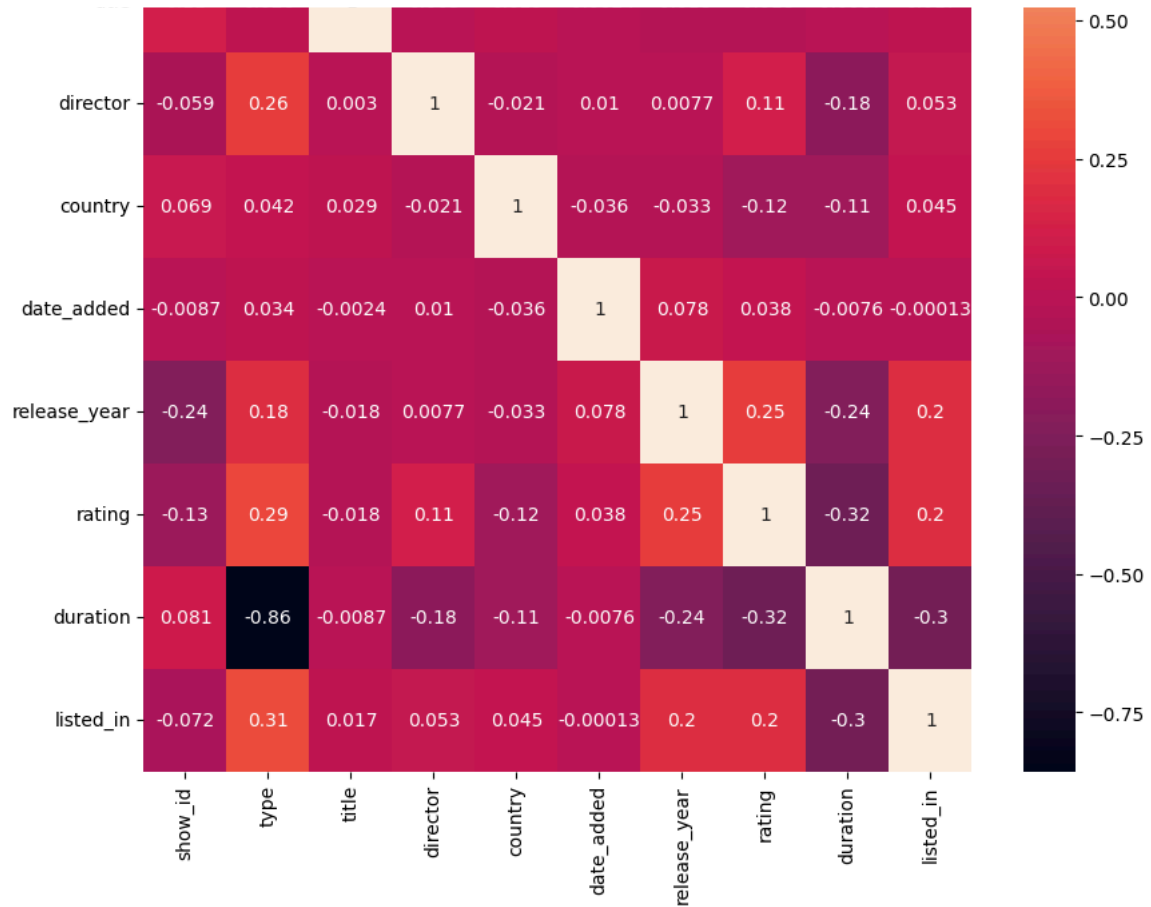
```
data['duration'] = data['duration'].str.replace(' min', '').str.replace(' Season', '').str.replace(' Seasons', '')
data['duration'] = pd.to_numeric(data['duration'], errors='coerce')
display(data.corr())
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
show_id	1.000000	-0.127189	0.108164	-0.059052	0.069123	-0.008719	-0.243790	-0.127187	0.080857	-0.072287
type	-0.127189	1.000000	0.026047	0.257602	0.042368	0.033827	0.182663	0.290778	-0.856306	0.306382
title	0.108164	0.026047	1.000000	0.003009	0.029428	-0.002438	-0.018077	-0.017783	-0.008710	0.016578
director	-0.059052	0.257602	0.003009	1.000000	-0.021002	0.010269	0.007737	0.113582	-0.183212	0.053042
country	0.069123	0.042368	0.029428	-0.021002	1.000000	-0.036372	-0.033043	-0.116072	-0.112713	0.044739
date_added	-0.008719	0.033827	-0.002438	0.010269	-0.036372	1.000000	0.077791	0.037562	-0.007595	-0.000128
release_year	-0.243790	0.182663	-0.018077	0.007737	-0.033043	0.077791	1.000000	0.250659	-0.237416	0.197505
rating	-0.127187	0.290778	-0.017783	0.113582	-0.116072	0.037562	0.250659	1.000000	-0.319352	0.201639
duration	0.080857	-0.856306	-0.008710	-0.183212	-0.112713	-0.007595	-0.237416	-0.319352	1.000000	-0.297716
listed_in	-0.072287	0.306382	0.016578	0.053042	0.044739	-0.000128	0.197505	0.201639	-0.297716	1.000000

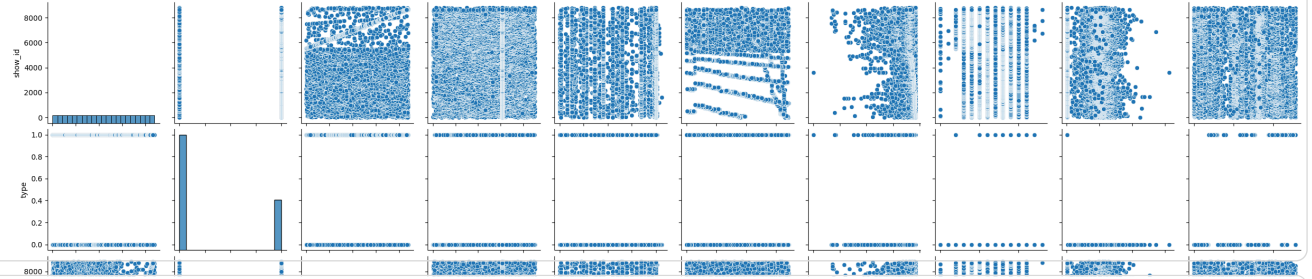
```
plt.figure(figsize=(10, 10))
sns.heatmap(data.corr(), annot=True)
plt.show()
```



```
sns.pairplot(data)
```

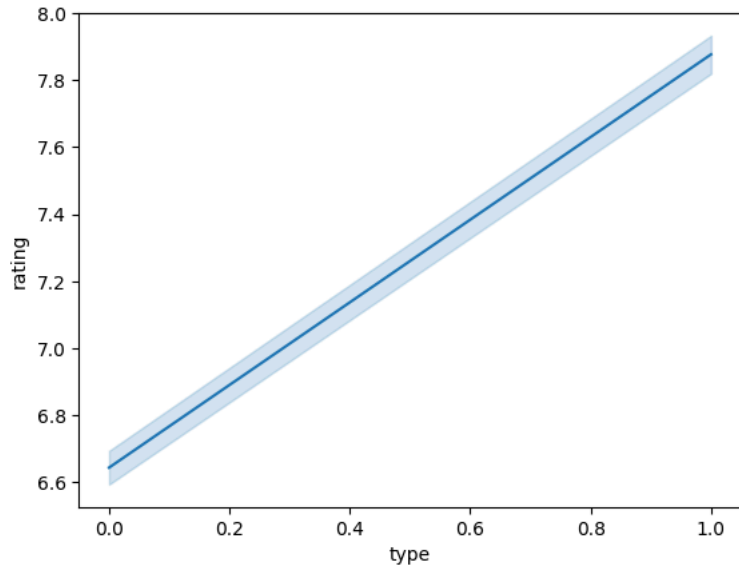


```
<seaborn.axisgrid.PairGrid at 0x7f1709ccac60>
```



```
sns.lineplot(x='type',y='rating',data=data)
```

```
<Axes: xlabel='type', ylabel='rating'>
```



```
sns.lineplot(x='rating',y='duration',data=data)
```

```
<Axes: xlabel='rating', ylabel='duration'>
```

