

```
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	ra
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	F
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	T
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	T
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	T
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	T
...	...	...	...	...	...	...	...	...
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	T
8786	s8798	TV Show	Zak Storm	Not Given	United States	9/13/2018	2016	-
8787	s8801	TV Show	Zindagi Gulzar Hai	Not Given	Pakistan	12/15/2016	2012	T
8788	s8784	TV Show	Yoko	Not Given	Pakistan	6/23/2018	2016	-
8789	s8786	TV Show	YOM	Not Given	Pakistan	6/7/2018	2016	-

8790 rows × 10 columns

Next steps:

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

```
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
0	1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-
1	3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-M
2	6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-M
3	14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-P
4	8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-M

Next steps:

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

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

country

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['type'].value\_counts()

count	
type	
Movie	6126
TV Show	2664

**dtype:** int64

data.loc[data['release\_year']==2020]

	show_id	type	title	director	country	date_added	release_year	rating
0	1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	F
25	17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	Pedro de Echave García, Pablo Azorín Williams	Not Given	9/22/2021	2020	T
		TV	Tayo and ...	...				

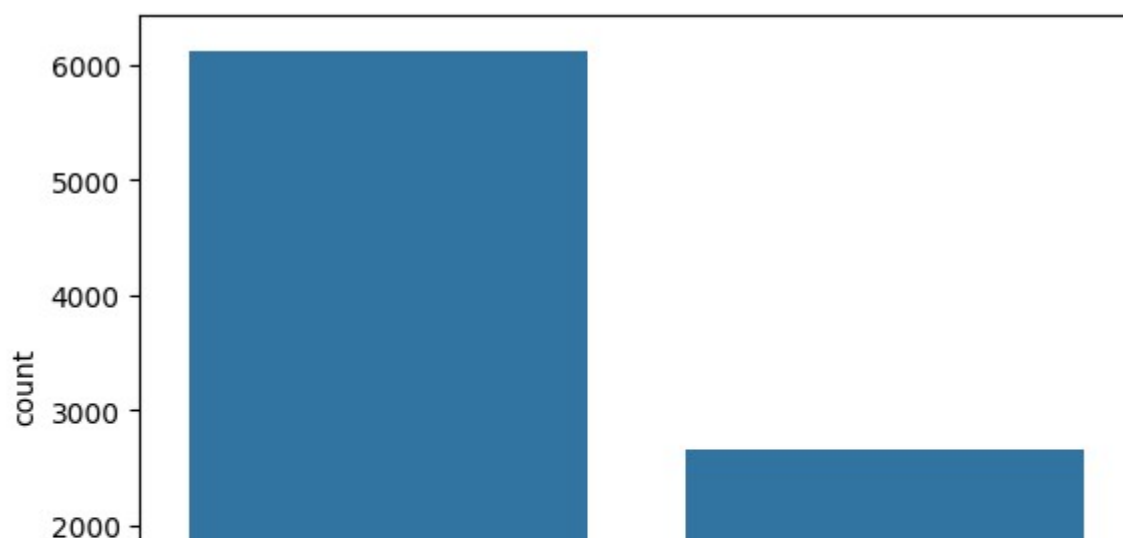
<b>36</b>	35	TV Show	Little Wizards	Not Given	Pakistan	9/17/2021	2020	
<b>47</b>	190	TV Show	Bread Barbershop	Not Given	Pakistan	8/28/2021	2020	
<b>58</b>	48	TV Show	The Smart Money Woman	Bunmi Ajakaiye	South Africa	9/16/2021	2020	T
...	...	...	...	...	...	...	...	
<b>7484</b>	3288	TV Show	Maradona in Mexico	Not Given	Argentina	11/13/2019	2020	T
<b>7517</b>	3370	TV Show	BoJack Horseman	Not Given	United States	10/25/2019	2020	T
<b>7537</b>	3434	TV Show	The Hook Up Plan	Not Given	France	10/11/2019	2020	T
<b>8685</b>	8126	TV Show	Super Wings	Not Given	United States	12/1/2020	2020	
<b>8687</b>	8133	TV Show	Surviving R. Kelly Part II: The Reckoning	Not Given	United States	4/13/2020	2020	T

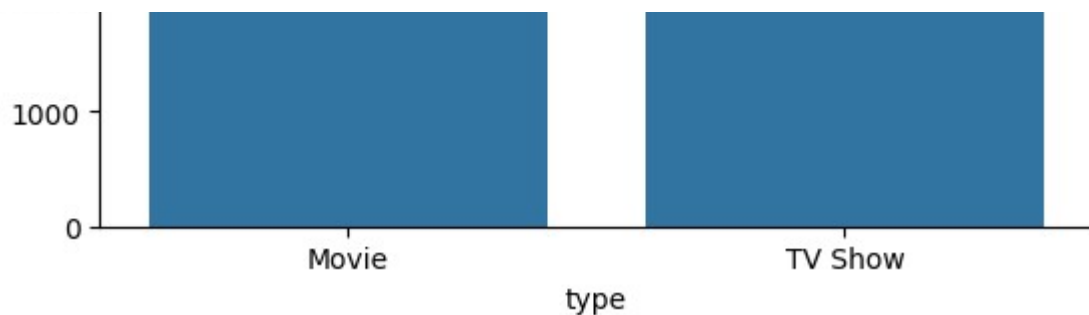
953 rows × 10 columns

data.iloc[100:110]									
	show_id	type	title	director	country	date_added	release_year	r	
<b>100</b>	370	TV Show	Myth & Mogul: John DeLorean	Not Given	Pakistan	7/30/2021	2021		
<b>101</b>	377	TV Show	Transformers: War for Cybertron: Kingdom	Not Given	Pakistan	7/29/2021	2021		
<b>102</b>	380	TV Show	Tattoo Redo	Not Given	Pakistan	7/28/2021	2021		

<b>103</b>	382	TV Show	The Snitch Cartel: Origins	Not Given	Pakistan	7/28/2021	2021
<b>104</b>	398	TV Show	Feels Like Ishq	Not Given	Pakistan	7/23/2021	2021
<b>105</b>	483	TV Show	How to Become a Tyrant	Not Given	Pakistan	7/9/2021	2021
<b>106</b>	82	Movie	Kate	Cedric Nicolas-Troyan	United States	9/10/2021	2021
<b>107</b>	85	Movie	Omo Ghetto: the Saga	JJC Skillz, Funke Akindele	Nigeria	9/10/2021	2020
<b>108</b>	87	Movie	Prey	Thomas Sieben	Germany	9/10/2021	2021
<b>109</b>	401	TV Show	Revelations: The Masters of the Universe: Reve...	Not Given	Pakistan	7/23/2021	2021

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





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

country	count
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
```

country	count
United States	3240
India	1057
United Kingdom	638
Pakistan	421

<b>Pakistan</b>	421
<b>Not Given</b>	287
<b>Canada</b>	271
<b>Japan</b>	259
<b>South Korea</b>	214
<b>France</b>	213
<b>Spain</b>	182

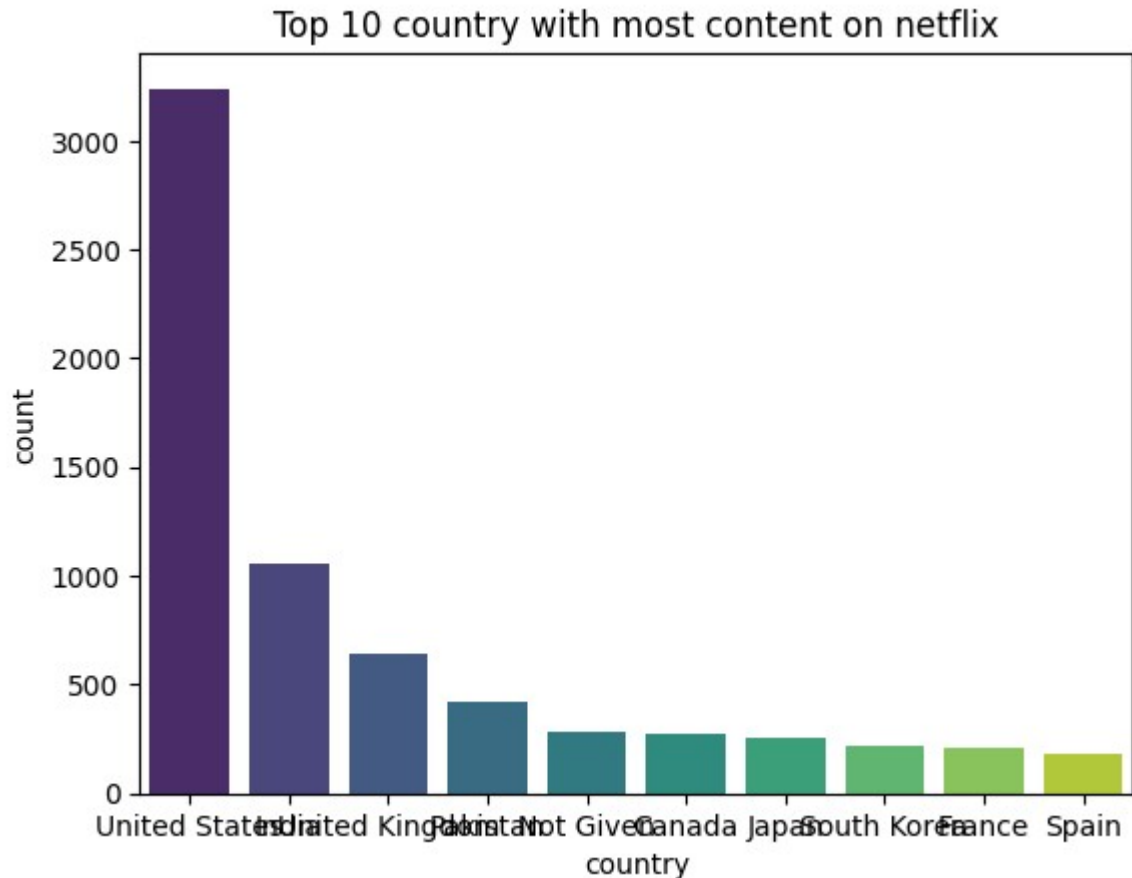
**dtype:** int64

```
sns.barplot(x=top_10.index,y=top_10,palette="viridis")
plt.xlabel('country')
plt.ylabel('count')
plt.title("Top 10 country with most content on netflix")
plt.show()
```

/tmp/ipython-input-651872578.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v

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



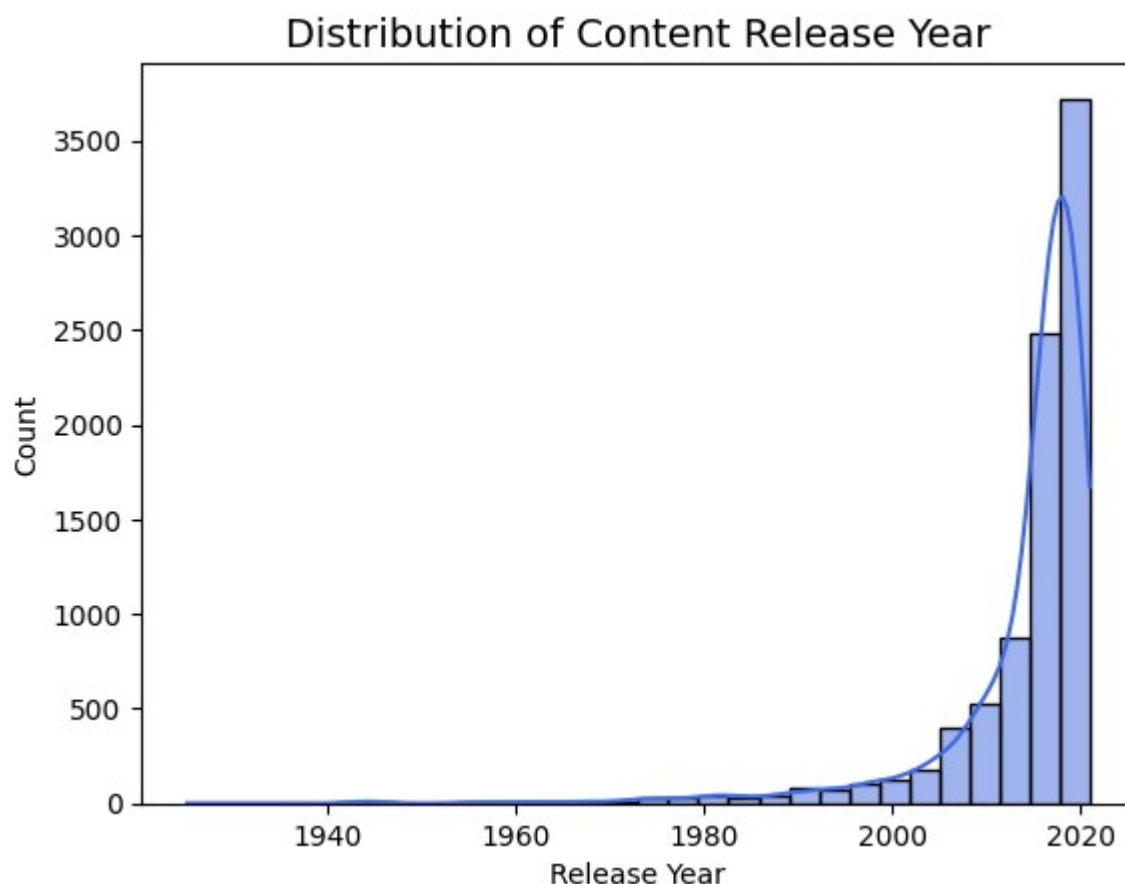
```
sns.histplot(data["release_year"], bins=30, kde=True, color="royalblue")

plt.title("Distribution of Content Release Year", fontsize=14)

plt.xlabel("Release Year")

plt.ylabel("Count")

plt.show()
```



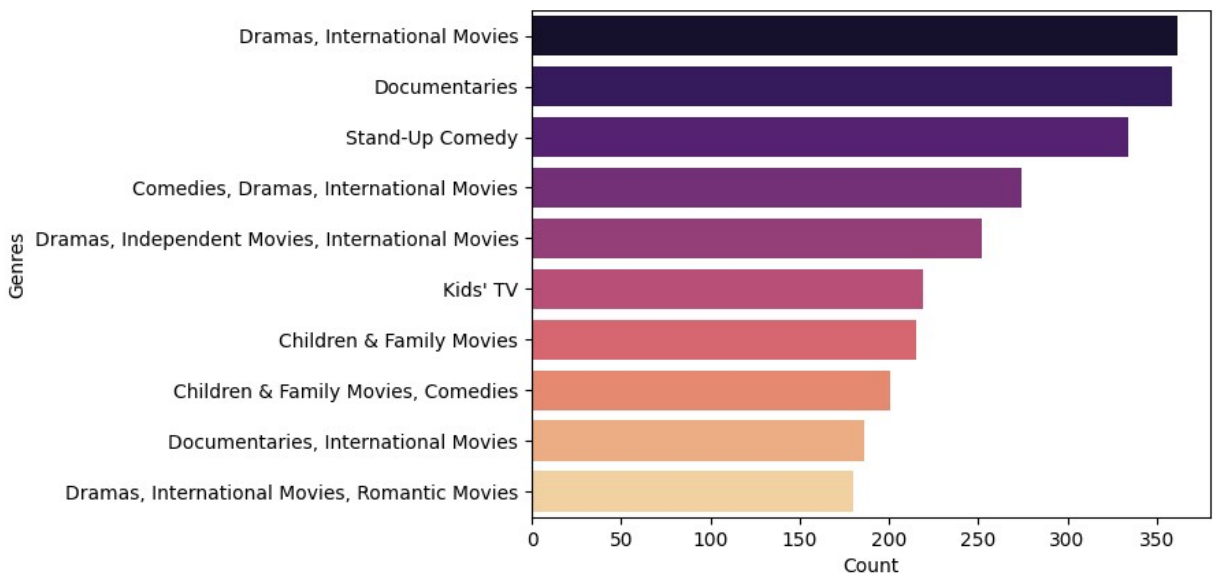
```
top_genres = data[ "listed_in"]. value_counts().head(10)
sns.barplot(x=top_genres.values, y=top_genres. index, palette="magma")
plt.title("Top 10 Most Common Genres on Netflix"
, fontsize=14)
plt.xlabel( "Count")
plt.ylabel ( "Genres")
plt.show()
```

/tmp/ipython-input-474789517.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v

```
sns.barplot(x=top_genres.values, y=top_genres. index, palette="magma")
Top 10 Most Common Genres on Netflix
```





```
rating_counts = data["rating"].value_counts()

sns.barplot(x=rating_counts.index, y=rating_counts.values, palette="pastel")

plt.title("Distribution of Content Ratings on Netflix", fontsize=14)

plt.xlabel("Rating")

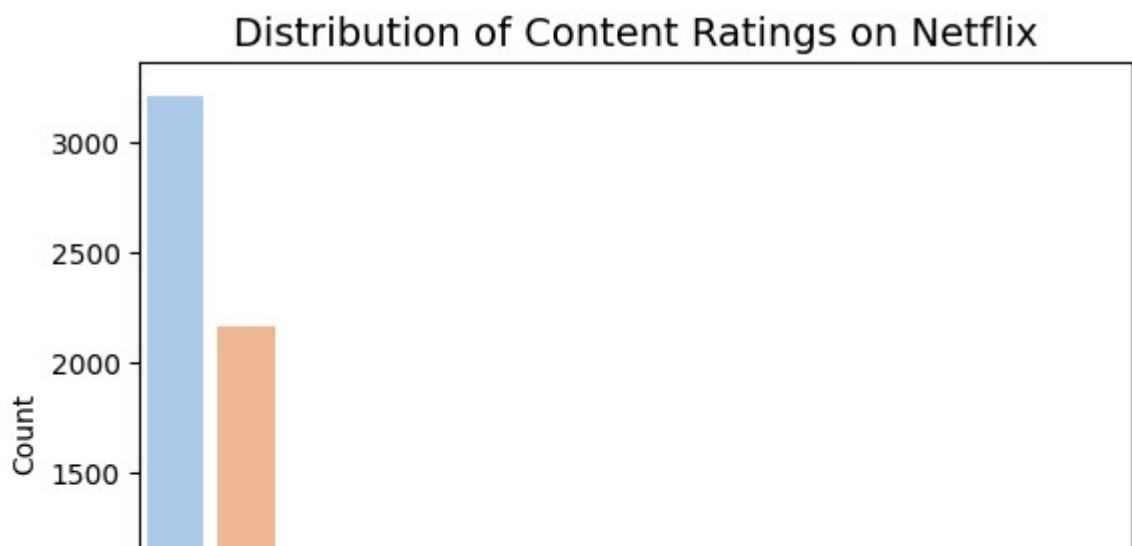
plt.ylabel("Count")
plt.xticks(rotation=45)

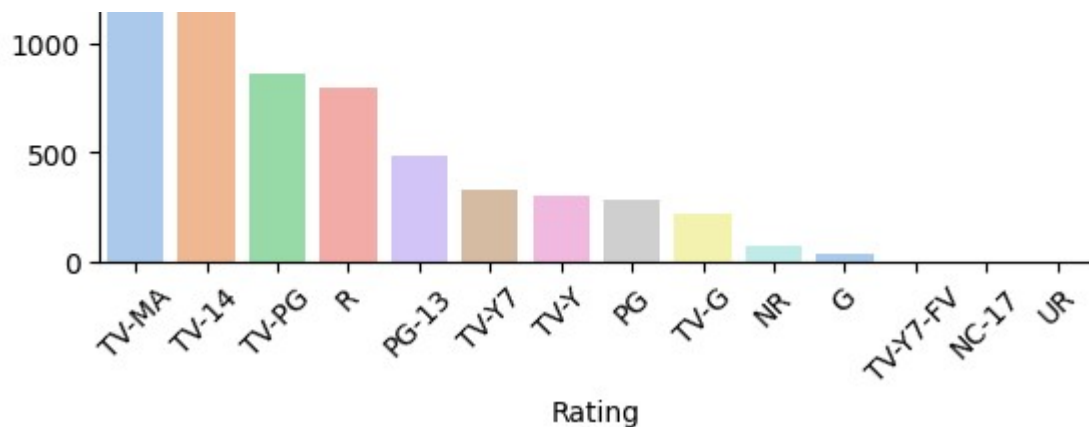
plt.show()
```

/tmp/ipython-input-2639363730.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v

```
sns.barplot(x=rating_counts.index, y=rating_counts.values, palette="pastel")
```





```

movies_df = data[data["type"] == "Movie"].copy()

movies_df["duration"] = movies_df["duration"].str.replace(" min", "").astype(int)

plt.figure(figsize=(14, 6))

sns.histplot(movies_df["duration"], bins=30, kde=True, color="royalblue")

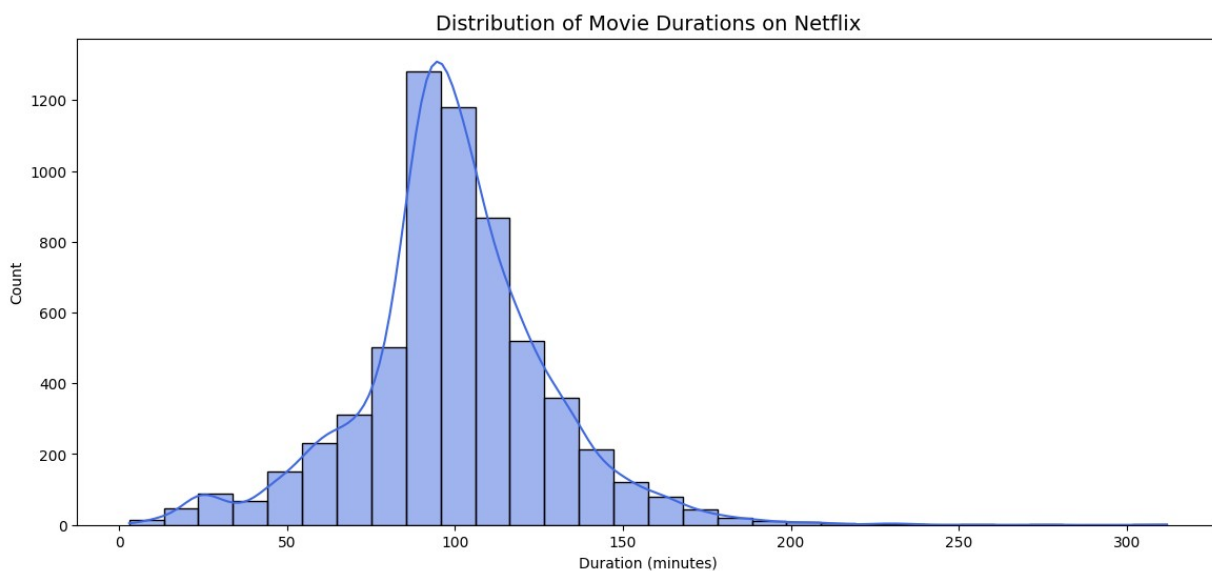
plt.title("Distribution of Movie Durations on Netflix", fontsize=14)

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", fontsi

plt.xlabel("Rating")

plt.ylabel("Duration (minutes)")

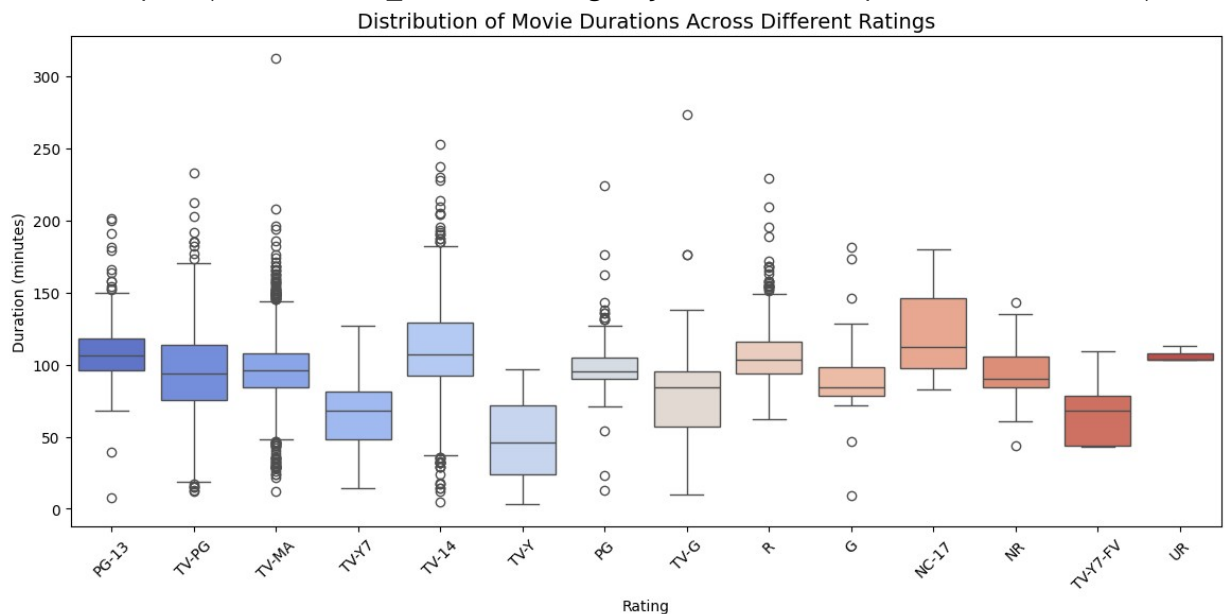
plt.xticks (rotation=45)

plt.show()
```

/tmp/ipython-input-1700444524.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v

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



```
data.columns
```

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

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

```
data.head()
```

```
show_id  type  title  director  country  date_added  release_year  rating  dur
```

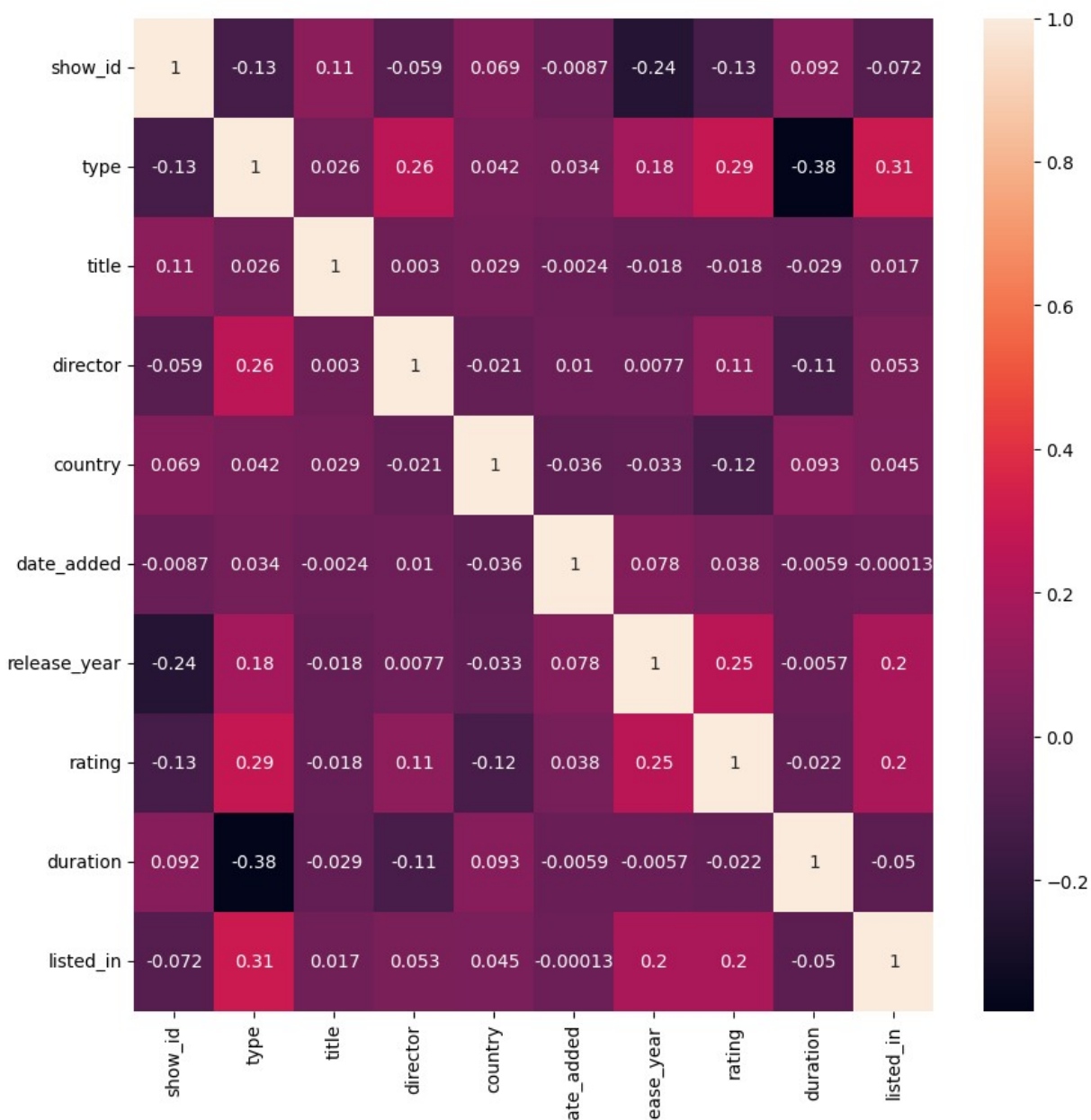
<b>0</b>	0	0	1972	2294	80	1656	2020	4
<b>1</b>	2222	1	2644	2104	20	1651	2021	8
<b>2</b>	5551	1	4558	2865	80	1651	2021	8
<b>3</b>	445	0	1657	627	6	1642	2021	9
<b>4</b>	7761	0	6043	1503	80	1651	1993	8

Next steps:

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

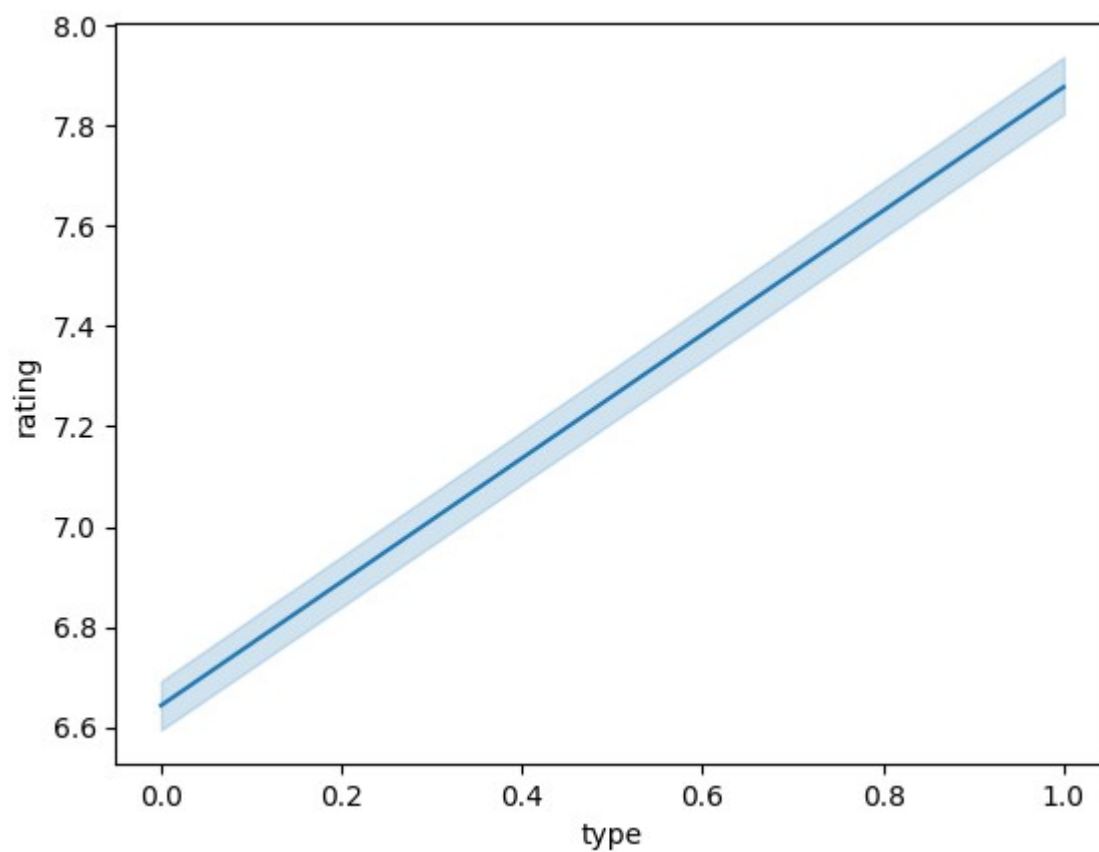
```
plt.figure(figsize=(10,10))
sns.heatmap(data.corr(),annot=True)
```

&lt;Axes: &gt;



```
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'>

