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

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

```
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.isnull().sum()/len(data)*100
```

	0
show_id	0.0
type	0.0
title	0.0
director	0.0
country	0.0
date_added	0.0
release_year	0.0
rating	0.0
duration	0.0
listed_in	0.0

dtype: float64

```
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	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

```
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

```
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()
```

type	count
Movie	6126
TV Show	2664

dtype: int64

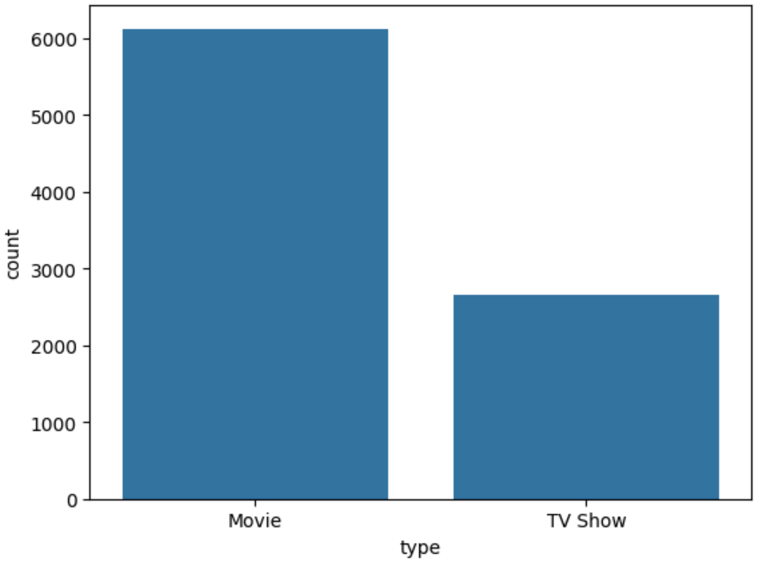
```
data.loc[data["release_year"] == "2020"]
```

show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in

```
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 ...
105	483	TV Show	How to Become a Tyrant	Not Given	Pakistan	7/9/2021	2021	TV-MA	1 Season	Docuseries
106	82	Movie	Kate	Cedric Nicolas-Troyan	United States	9/10/2021	2021	R	106 min	Action & Adventure
107	85	Movie	Omo Ghetto: the Saga	JJC Skillz, Funke	Nigeria	9/10/2021	2020	TV-MA	147 min	Action & Adventure, Comedies, Dramas

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



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

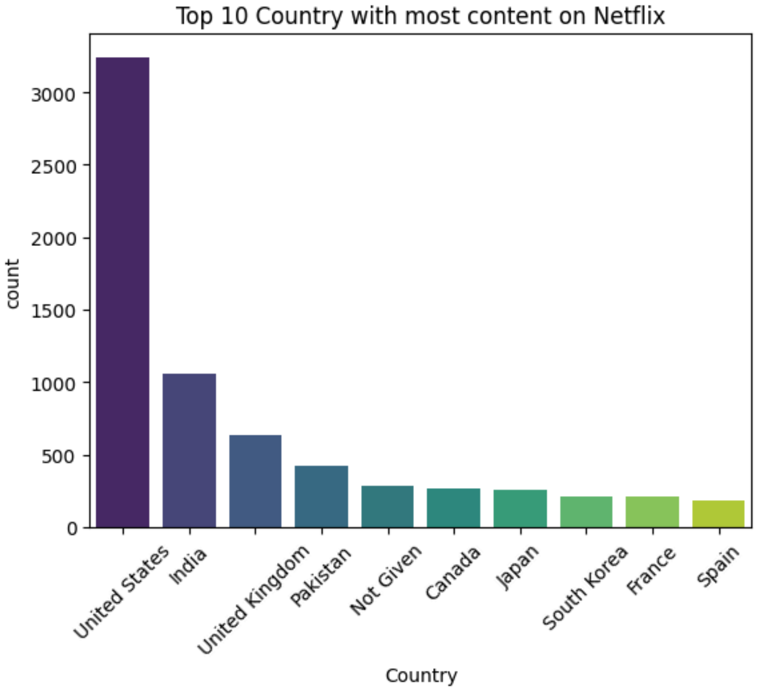
```
                  count
        country
  United States    3240
          India    1057
 United Kingdom     638
       Pakistan     421
      Not Given     287
            ...     ...
     Luxembourg       1
        Senegal       1
        Belarus       1
    Puerto Rico       1
         Cyprus 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

86 rows × 1 columns

dtype: int64
```

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

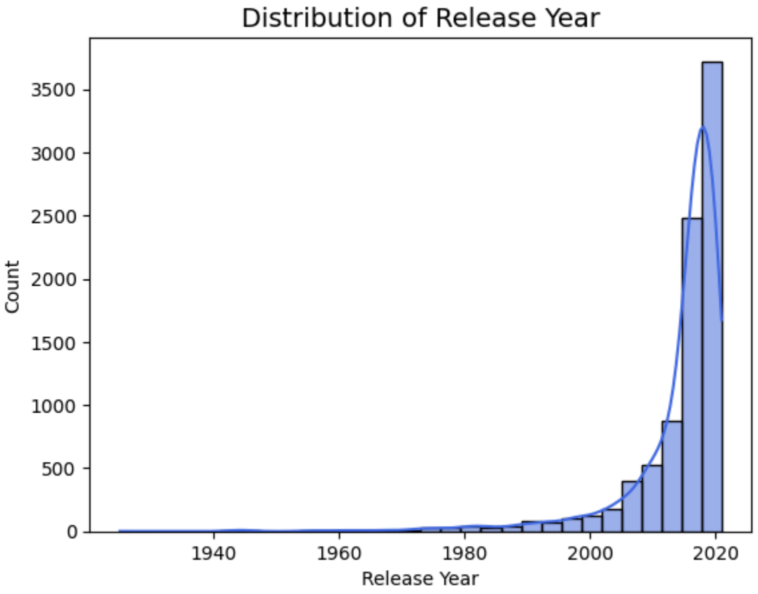
```
                  count
        country
  United States    3240
          India    1057
 United Kingdom     638
       Pakistan     421
      Not Given     287
         Canada     271
          Japan     259
    South Korea     214
         France     213
          Spain     182

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 content on Netflix")
plt.xticks(rotation=45)
plt.show()
```

```
/tmp/ipython-input-2200090005.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 se

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

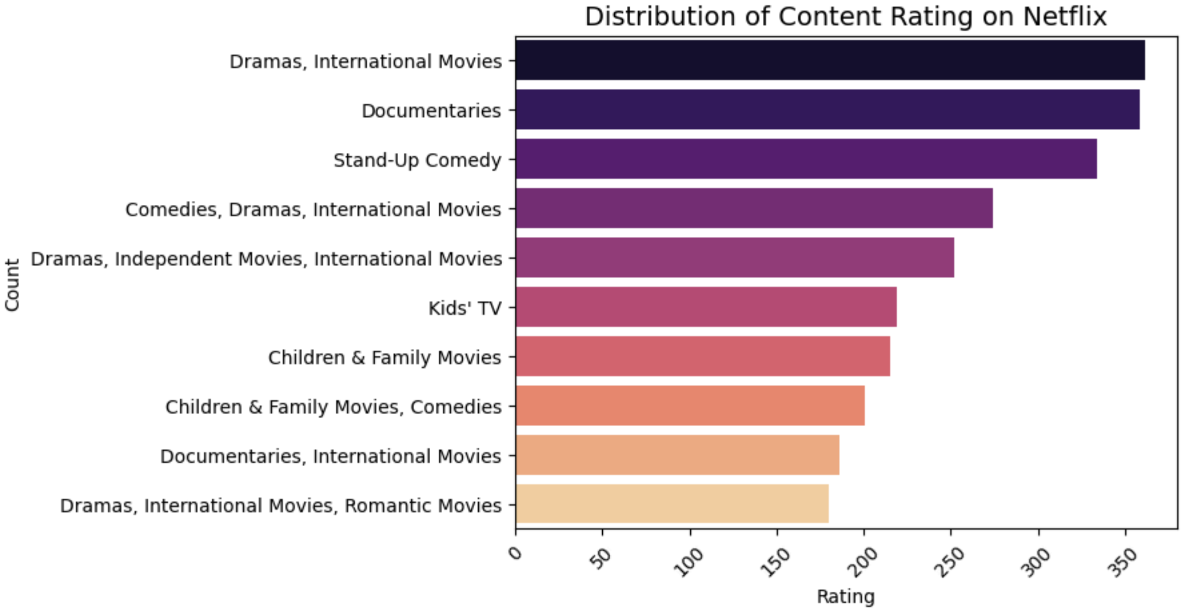


```
sns.histplot(data['release_year'],bins = 30, kde= True, color="royalblue")
plt.xlabel('Release Year')
plt.ylabel("Count")
plt.title("Distribution of Release Year",fontsize = 14)
plt.show()
```



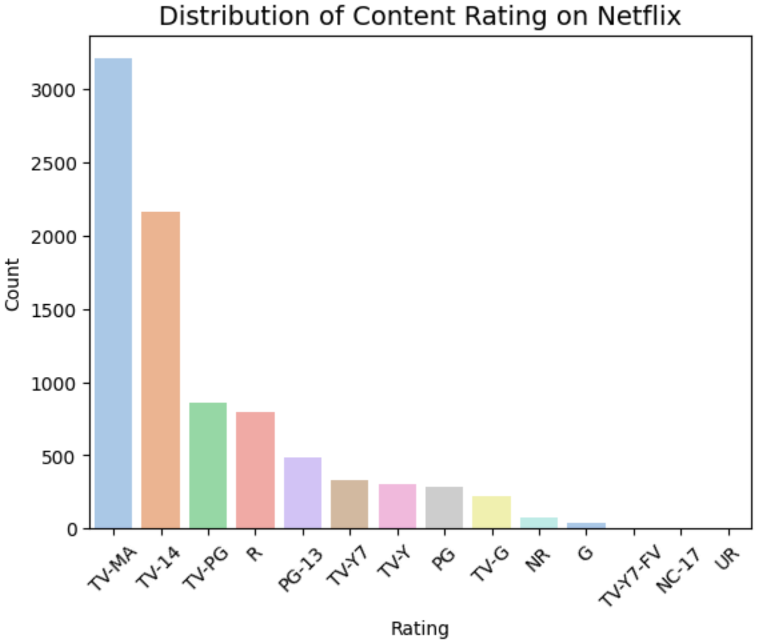
```
top_genres = data ["listed_in"].value_counts().head(10)
sns.barplot(x=top_genres.values, y = top_genres.index, palette= "magma")
plt.xlabel("Rating")
plt.ylabel("Count")
plt.title("Distribution of Content Rating on Netflix",fontsize =14)
plt.xticks(rotation=45)
plt.show()
```

```
/tmp/ipython-input-3508971630.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 se

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

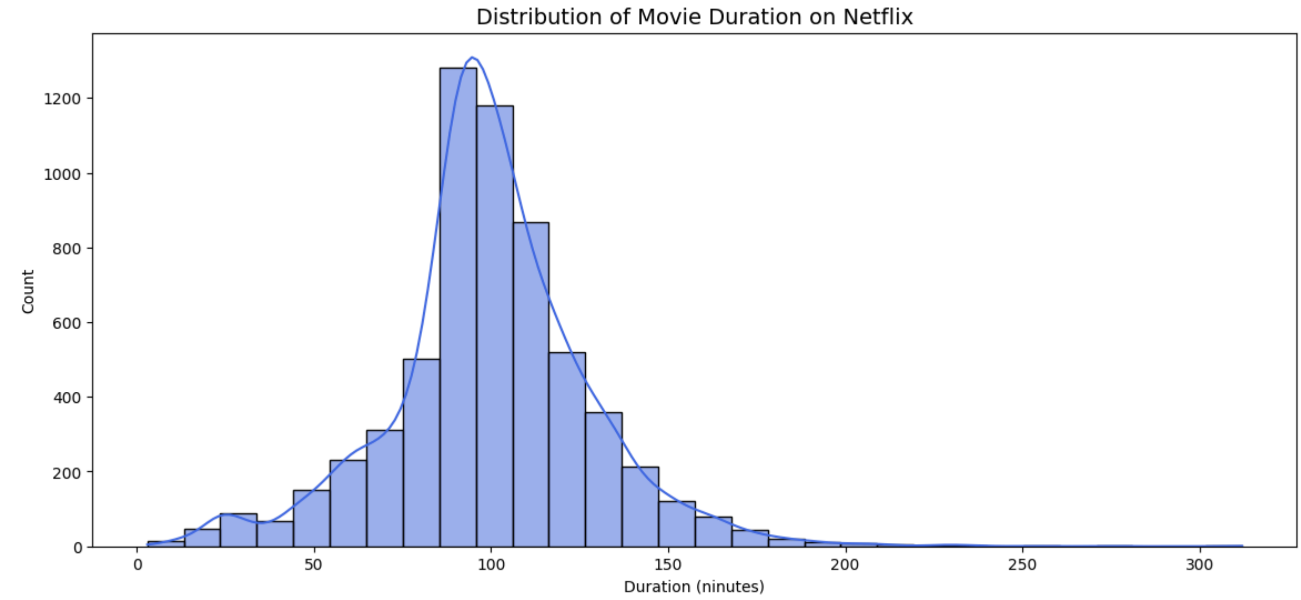


```
rating_counts = data["rating"].value_counts()
sns.barplot(x=rating_counts.index, y=rating_counts.values, palette="pastel")
plt.xlabel("Rating")
plt.ylabel("Count")
plt.title("Distribution of Content Rating on Netflix",fontsize =14)
plt.xticks(rotation=45)
plt.show()
```

```
/tmp/ipython-input-2187259553.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 se

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

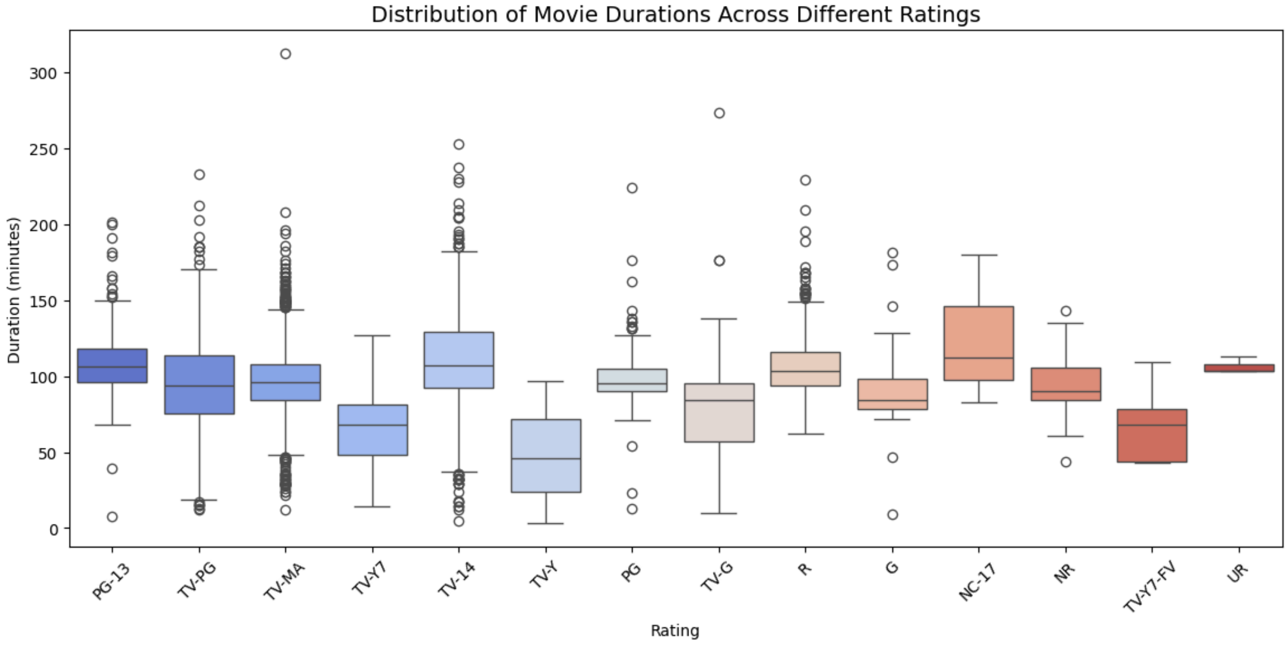


```
movies_df = data[data["type"] == "Movie"].copy()
movies_df["duration"] = movies_df["duration"].str.replace(" min","").astype(float)

plt.figure(figsize =(14, 6))
sns.histplot(movies_df["duration"], bins =30, kde = True, color ="royalblue")
plt.xlabel("Duration (ninutes)")
plt.ylabel("Count")
plt.title("Distribution of Movie Duration on Netflix", fontsize =14)
plt.show()
```



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

```
/tmp/ipython-input-3835333525.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 se

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



```
data.columns
```

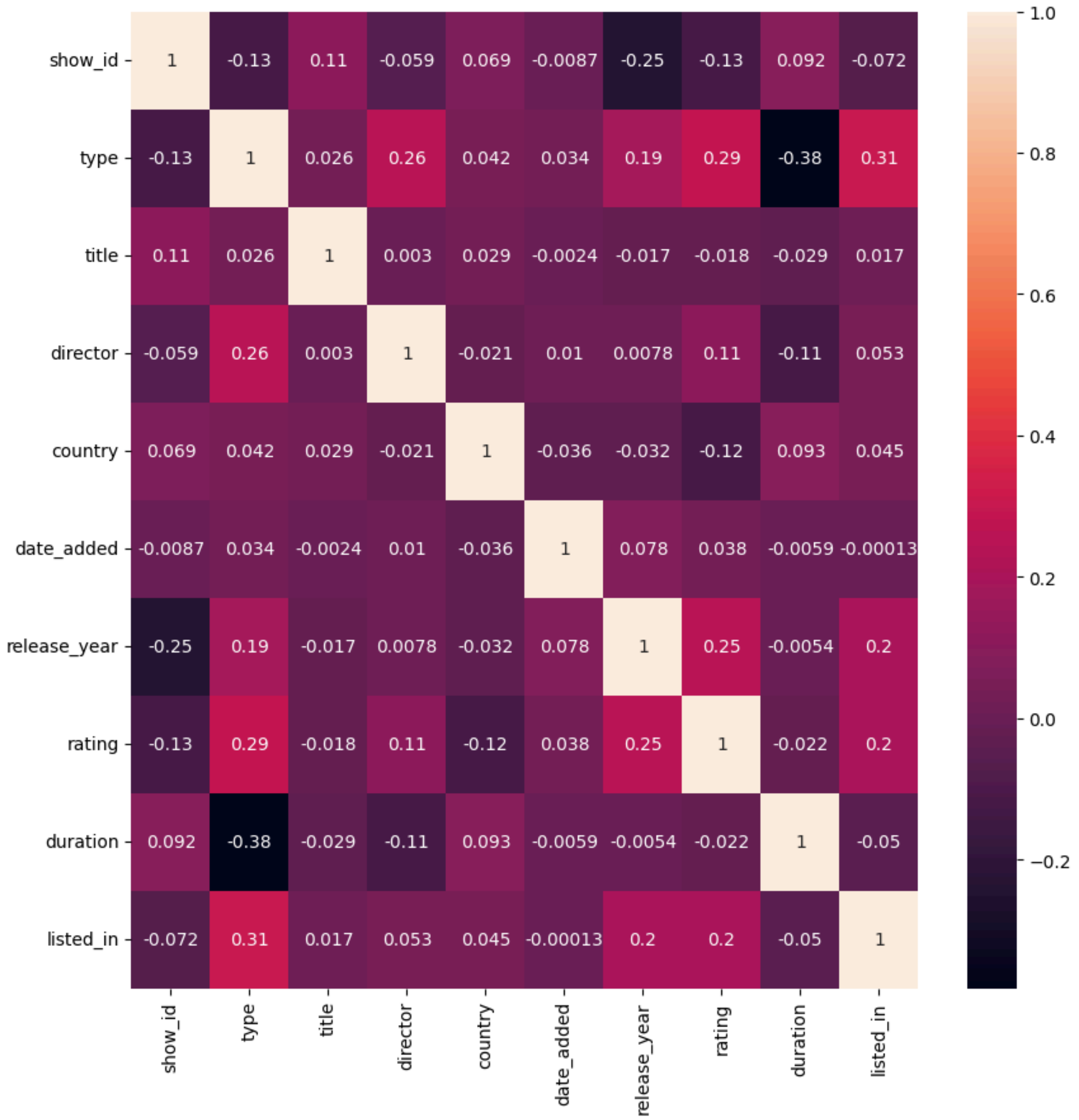
```
Index(['show_id', 'type', 'title', 'director', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
      dtype='object')
```

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

```
data.corr()
```

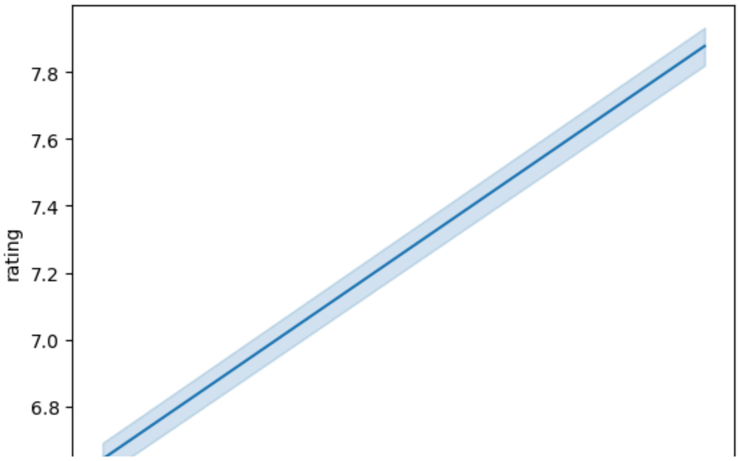
	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
show_id	1.000000	-0.127189	0.108164	-0.059052	0.069123	-0.008719	-0.245564	-0.127187	0.092445	-0.072287
type	-0.127189	1.000000	0.026047	0.257602	0.042368	0.033827	0.185049	0.290778	-0.381682	0.306382
title	0.108164	0.026047	1.000000	0.003009	0.029428	-0.002438	-0.017079	-0.017783	-0.029480	0.016578
director	-0.059052	0.257602	0.003009	1.000000	-0.021002	0.010269	0.007795	0.113582	-0.113439	0.053042
country	0.069123	0.042368	0.029428	-0.021002	1.000000	-0.036372	-0.032319	-0.116072	0.092700	0.044739
date_added	-0.008719	0.033827	-0.002438	0.010269	-0.036372	1.000000	0.078465	0.037562	-0.005946	-0.000128
release_year	-0.245564	0.185049	-0.017079	0.007795	-0.032319	0.078465	1.000000	0.254172	-0.005351	0.200234
rating	-0.127187	0.290778	-0.017783	0.113582	-0.116072	0.037562	0.254172	1.000000	-0.021513	0.201639
duration	0.092445	-0.381682	-0.029480	-0.113439	0.092700	-0.005946	-0.005351	-0.021513	1.000000	-0.049587
listed_in	-0.072287	0.306382	0.016578	0.053042	0.044739	-0.000128	0.200234	0.201639	-0.049587	1.000000

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



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

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


rating