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

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

```python
data
```

	show_id	type	title	director	country	date_added	release_year	rati
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-N
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-N
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-F
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-N
...	...	...	...	...	...	...	...	
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	TV-F
8786	s8798	TV Show	Zak Storm	Not Given	United States	9/13/2018	2016	TV-Y

Next steps: [Generate code with `data`] [New interactive sheet]

```python
data.head()
```

	show_id	type	title	director	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA

Next steps: (Generate code with `data`)   (New interactive sheet)

```
data.tail()
```

	show_id	type	title	director	country	date_added	release_year	rating	d
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	TV-PG	
8786	s8798	TV Show	Zak Storm	Not Given	United States	9/13/2018	2016	TV-Y7	
8787	s8801	TV Show	Zindagi Gulzar Hai	Not Given	Pakistan	12/15/2016	2012	TV-PG	
8788	s8784	TV Show	Yoko	Not Given	Pakistan	6/23/2018	2016	TV-Y	

```
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
```

```python
data.sample()
```

	show_id	type	title	director	country	date_added	release_year	rating	c
4570	8011	Movi	Blue	Les	C	1/1/2021	1000	PG-13	

```python
data.shape
```

```
(8790, 10)
```

```python
data.describe()
```

	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

```python
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	rati
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-N
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-N
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-F
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-N
...	...	...	...	...	...	...	...	
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	TV-F
8786	s8798	TV Show	Zak Storm	Not Given	United States	9/13/2018	2016	TV-\

Next steps: Generate code with `data` | New interactive sheet

```python
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-13
1	3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA
2	6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA

Next steps: (Generate code with `data`) (New interactive sheet)

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

	show_id	type	title	director	country	date_added	release_year	rati
0	1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-
25	17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	Pedro de Echave García, Pablo Azorín Williams	Not Given	9/22/2021	2020	TV-N
36	35	TV Show	Tayo and Little Wizards	Not Given	Pakistan	9/17/2021	2020	TV-
47	190	TV Show	Bread Barbershop	Not Given	Pakistan	8/28/2021	2020	TV
58	48	TV Show	The Smart Money Woman	Bunmi Ajakaiye	South Africa	9/16/2021	2020	TV-N
...	...	...	...	...	...	...	...	
7484	3288	TV Show	Maradona in Mexico	Not Given	Argentina	11/13/2019	2020	TV-N
7517	3370	TV Show	BoJack Horseman	Not Given	United States	10/25/2019	2020	TV-N
7537	3434	TV Show	The Hook Up Plan	Not Given	France	10/11/2019	2020	TV-N

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

	show_id	type	title	director	country	date_added	release_year	rati
100	370	TV Show	Myth & Mogul: John DeLorean	Not Given	Pakistan	7/30/2021	2021	TV
101	377	TV Show	Transformers: War for Cybertron: Kingdom	Not Given	Pakistan	7/29/2021	2021	TV-
102	380	TV Show	Tattoo Redo	Not Given	Pakistan	7/28/2021	2021	TV-
103	382	TV Show	The Snitch Cartel: Origins	Not Given	Pakistan	7/28/2021	2021	TV-
104	398	TV Show	Feels Like Ishq	Not Given	Pakistan	7/23/2021	2021	TV-
105	483	TV Show	How to Become a Tyrant	Not Given	Pakistan	7/9/2021	2021	TV-
106	82	Movie	Kate	Cedric Nicolas-	United	9/10/2021	2021	

```python
sns.countplot(x="type",data=data)
plt.show()
```
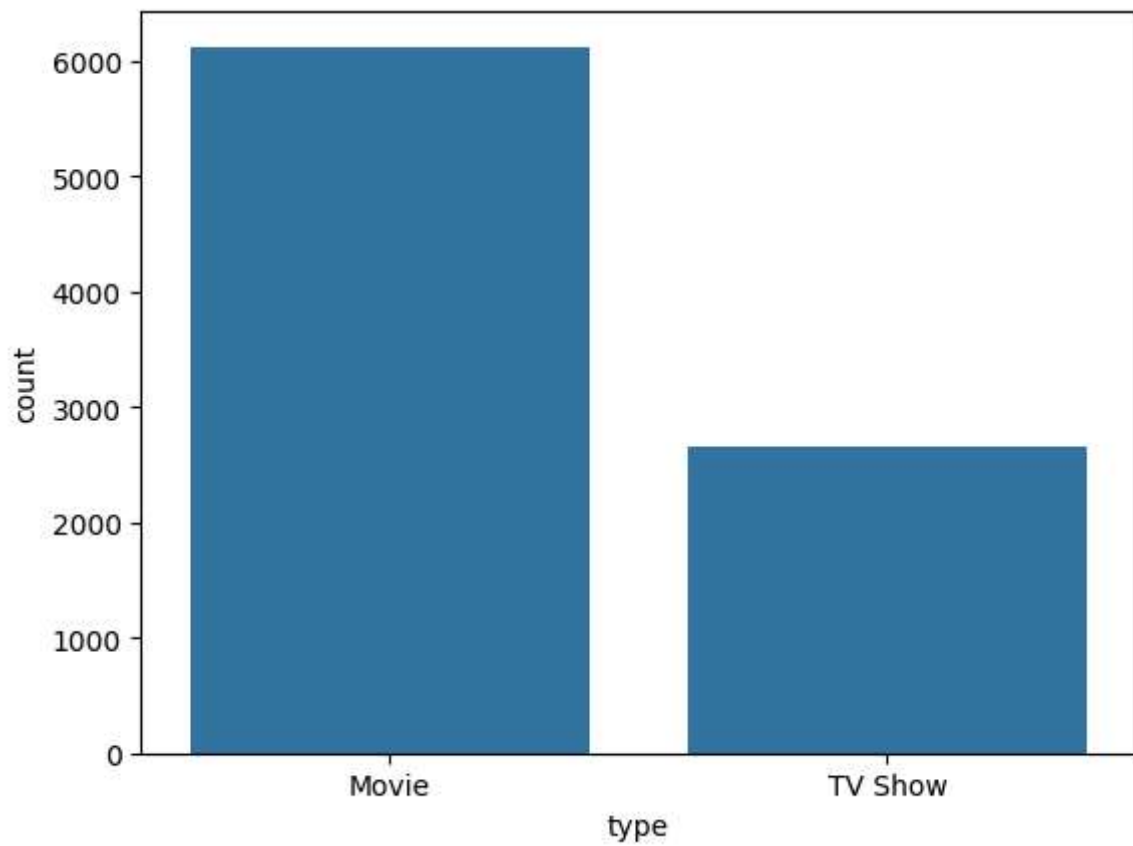
```python
data["country"].value_counts()
```

	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 content on netflix")
plt.xticks(rotation=45)
plt.show()
```
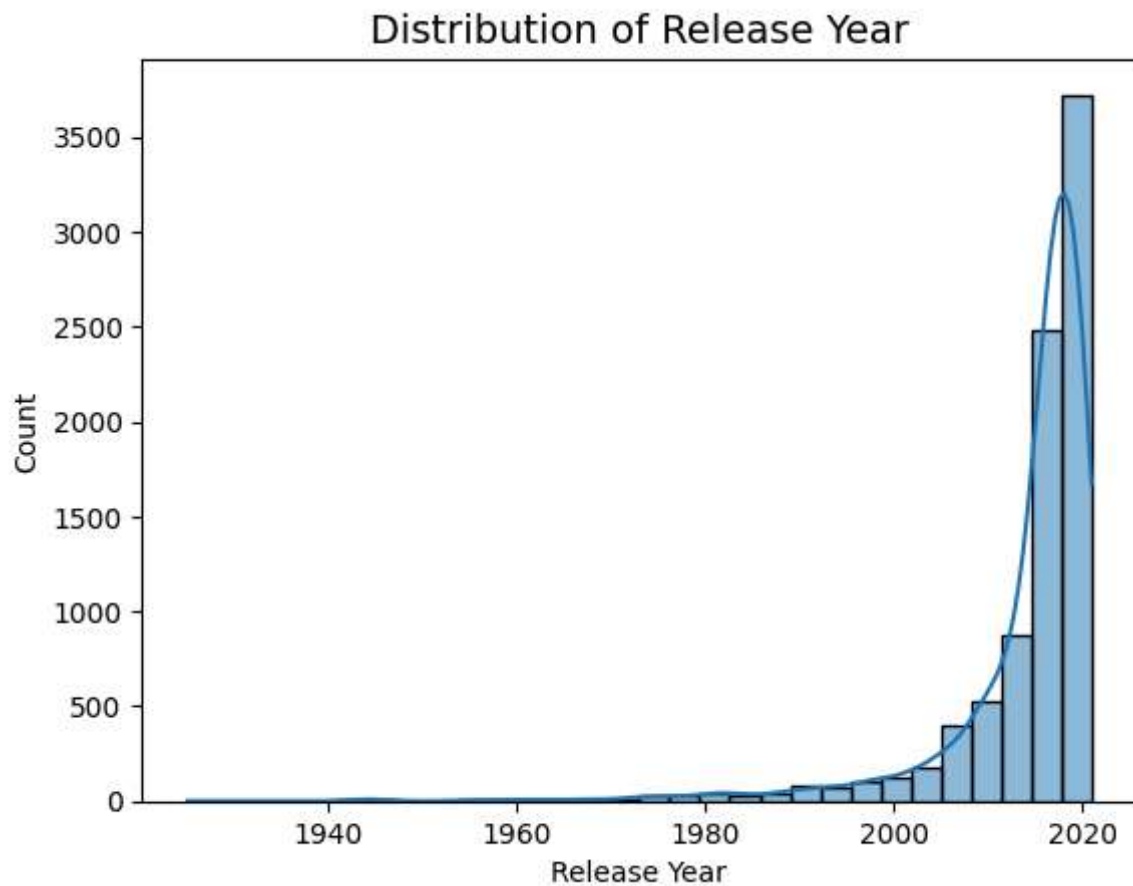
```
/tmp/ipython-input-449106027.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.1

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


Top 10 country with most content on netflix
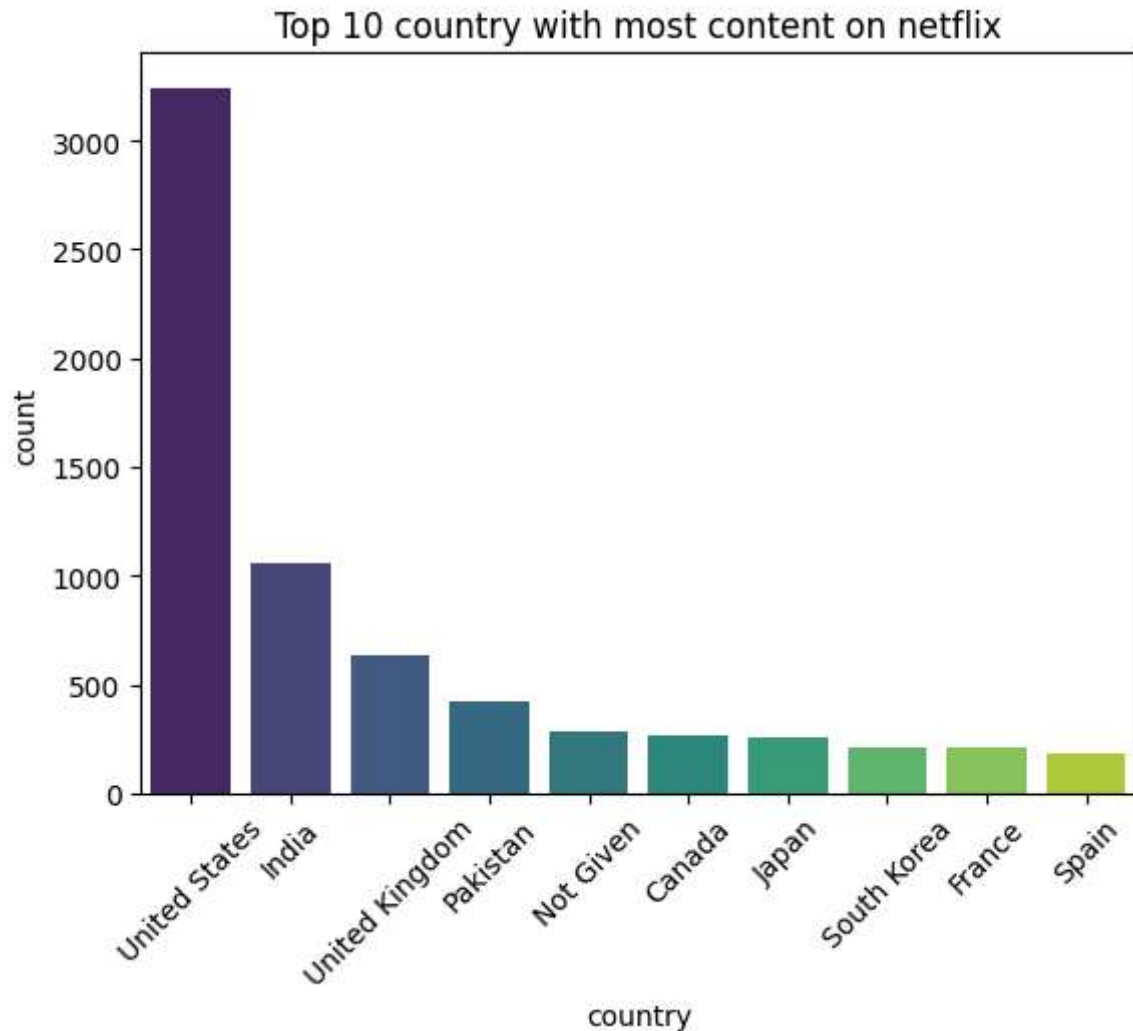
```
sns.histplot(data["release_year"],bins=30, kde=True).color="royalblue"
plt.title("Distribution of Release Year", fontsize=14)
plt.xlabel("Release Year")
plt.ylabel("Count")
plt.show()
```

Distribution of Release Year

```
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("Genre")
plt.ylabel("Count")
# plt.title("Top 10 Genres on Netflix")
plt.xticks(rotation=45)
plt.show()
```

```
/tmp/ipython-input-2553918280.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.1

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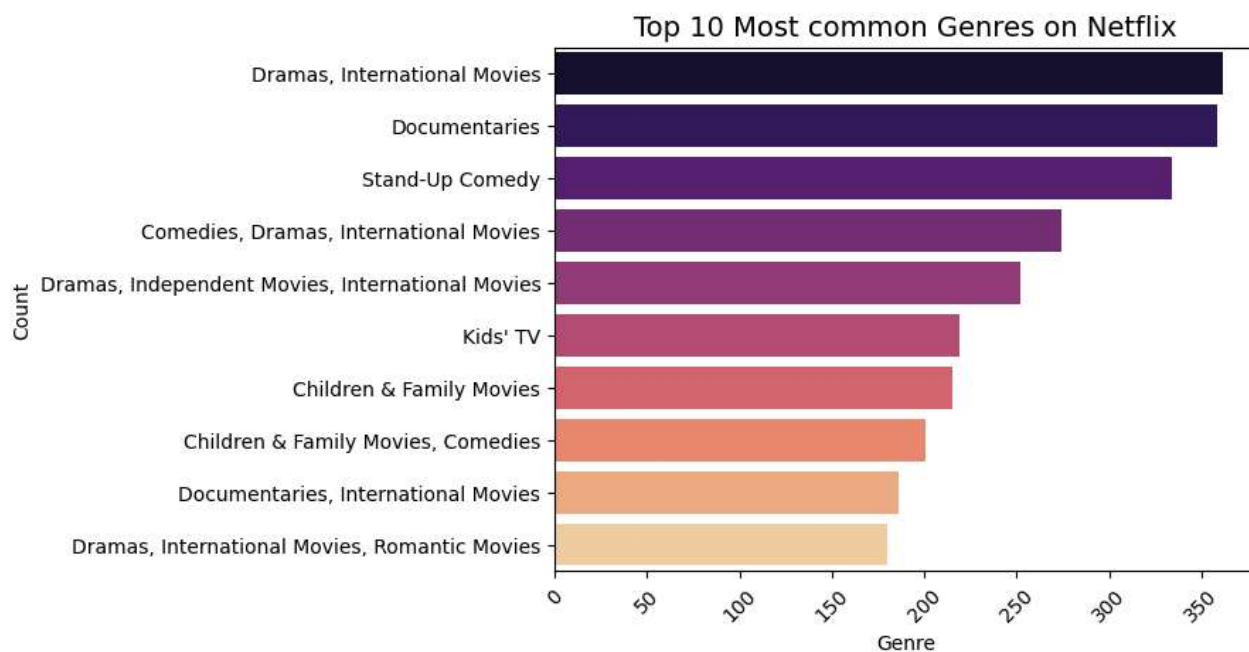


```
rating_counts=data["rating"].value_counts()

sns.barplot(x=rating_counts.index, y=rating_counts.values, palette="pastel")
plt.title("Distribution of content Ratings on Netflix", fontsize=14)
plt.xlabel("Rating")
plt.ylabel("Count")
plt.xticks(rotation=45)
```

```
/tmp/ipython-input-885864341.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.1

  sns.barplot(x=rating_counts.index, y=rating_counts.values, palette="pastel")
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13],
 [Text(0, 0, 'TV-MA'),
  Text(1, 0, 'TV-14'),
  Text(2, 0, 'TV-PG'),
  Text(3, 0, 'R'),
  Text(4, 0, 'PG-13'),
  Text(5, 0, 'TV-Y7'),
  Text(6, 0, 'TV-Y'),
  Text(7, 0, 'PG'),
  Text(8, 0, 'TV-G'),
  Text(9, 0, 'NR'),
  Text(10, 0, 'G'),
  Text(11, 0, 'TV-Y7-FV'),
  Text(12, 0, 'NC-17'),
  Text(13, 0, 'UR')])
```


Distribution of content Ratings on Netflix
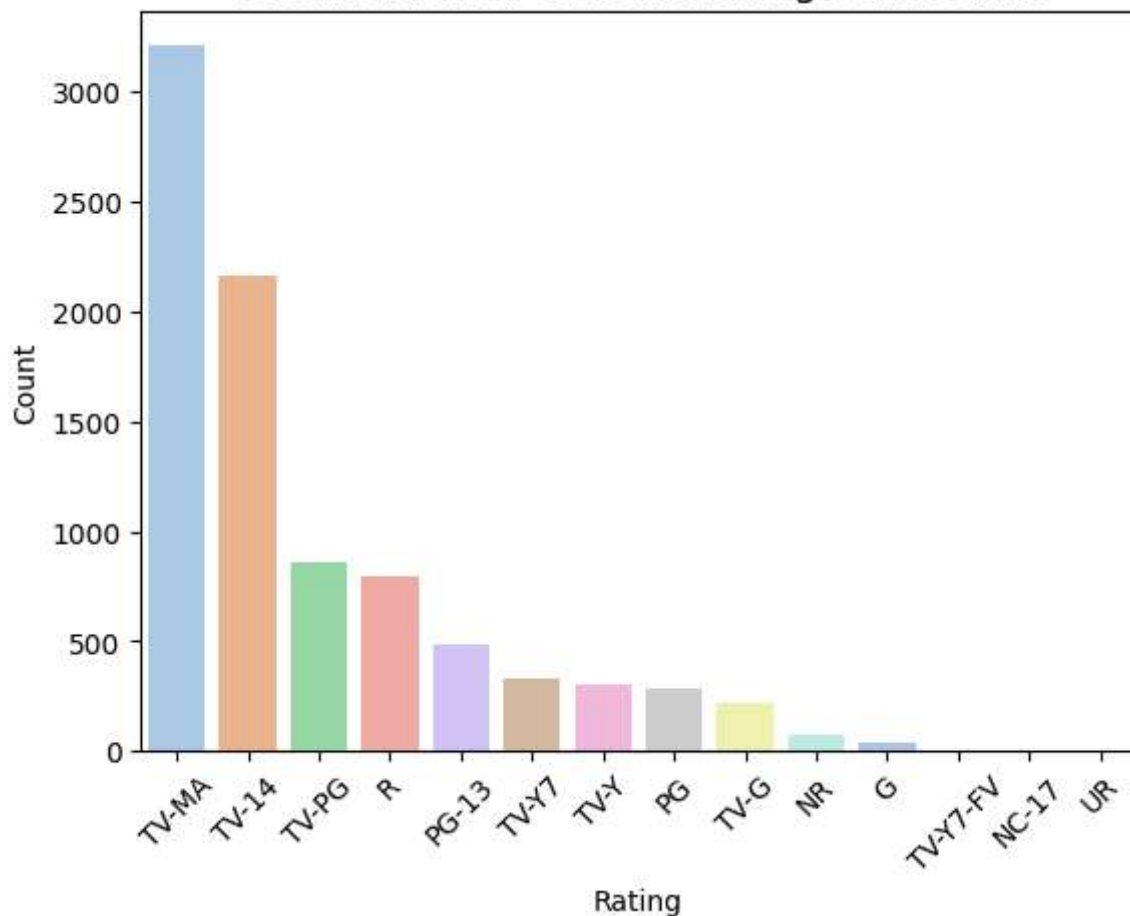
```
movies_df=data[data["type"] == "Movie"].copy()

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

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


Destribution of Movie Durations on Netflix