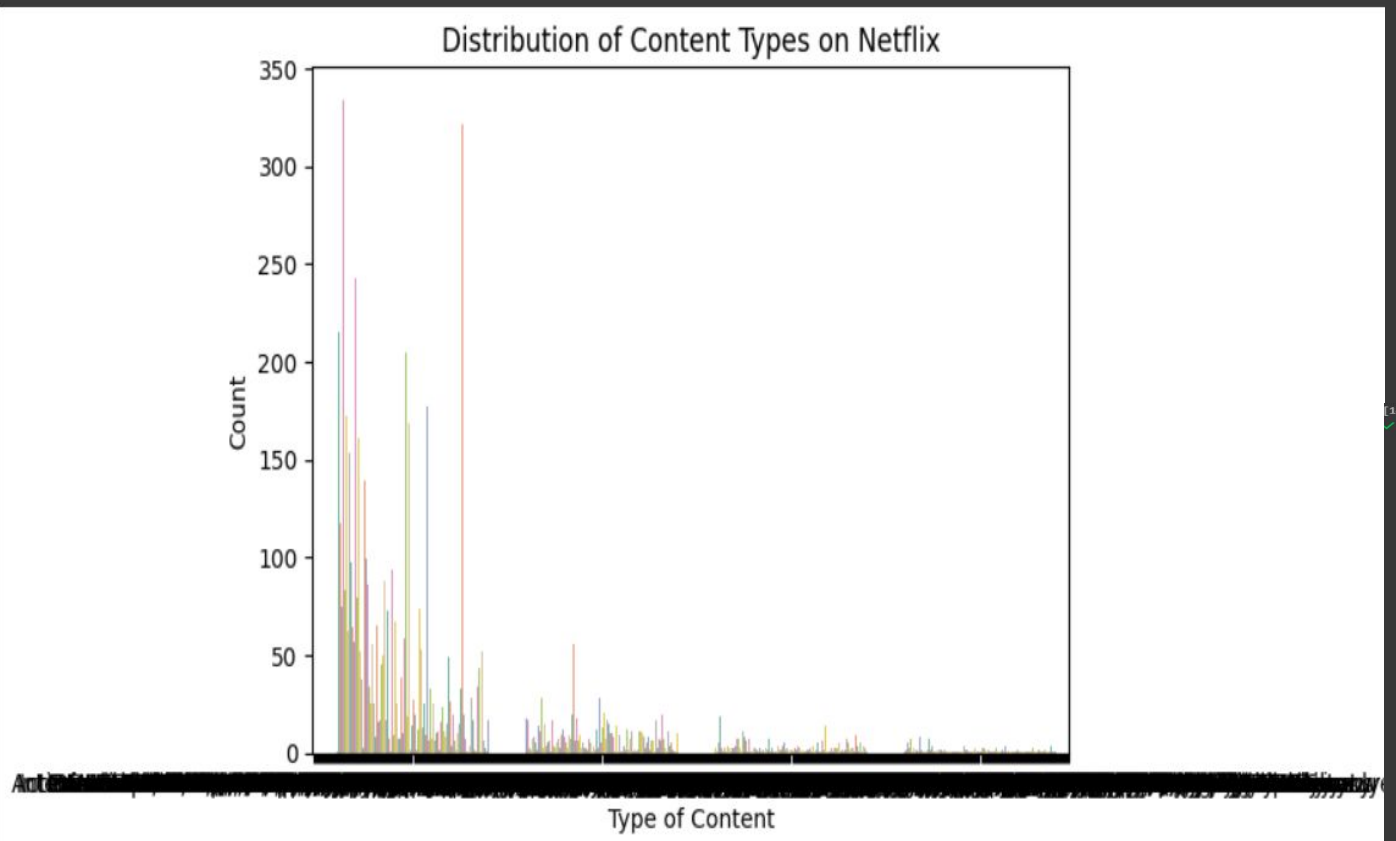
RESULTS 1

```
#Content type distribution
sns.countplot(data=df, x='Type', palette='Set2')
plt.title("Distribution of Content Types on Netflix")
plt.xlabel("Type of Content")
plt.ylabel("Count")
plt.show()
```



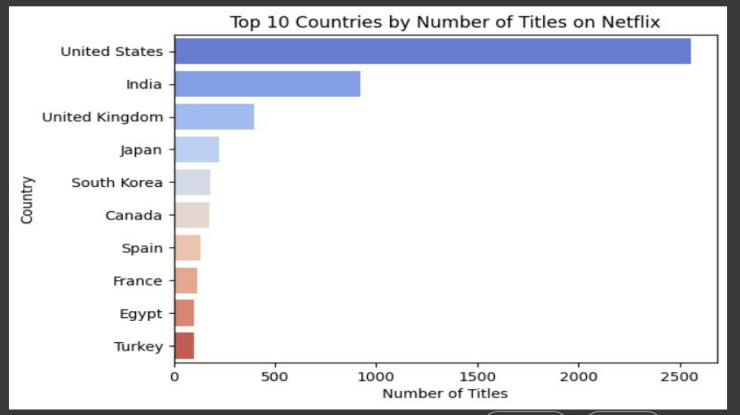

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

```
# Load the dataset
file_path = "Netflix Dataset.csv"
df = pd.read_csv(file_path)
```

```
#Basic Overview
print("🔹 Dataset Overview:\n")
print(df.info())
print("\n🔹 First 5 Rows:\n")
print(df.head())
print("\n🔹 Missing Values:\n")
print(df.isnull().sum())
print("\n🔹 Duplicate Rows:", df.duplicated().sum())
```

```
🔹 Dataset Overview:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7789 entries, 0 to 7788
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Show_Id       7789 non-null   object
```

```
sns.barplot(x=top_countries.values, y=top_countries.index, palette='coolwarm')
plt.title("Top 10 Countries by Number of Titles on Netflix")
plt.xlabel("Number of Titles")
plt.ylabel("Country")
plt.show()
```



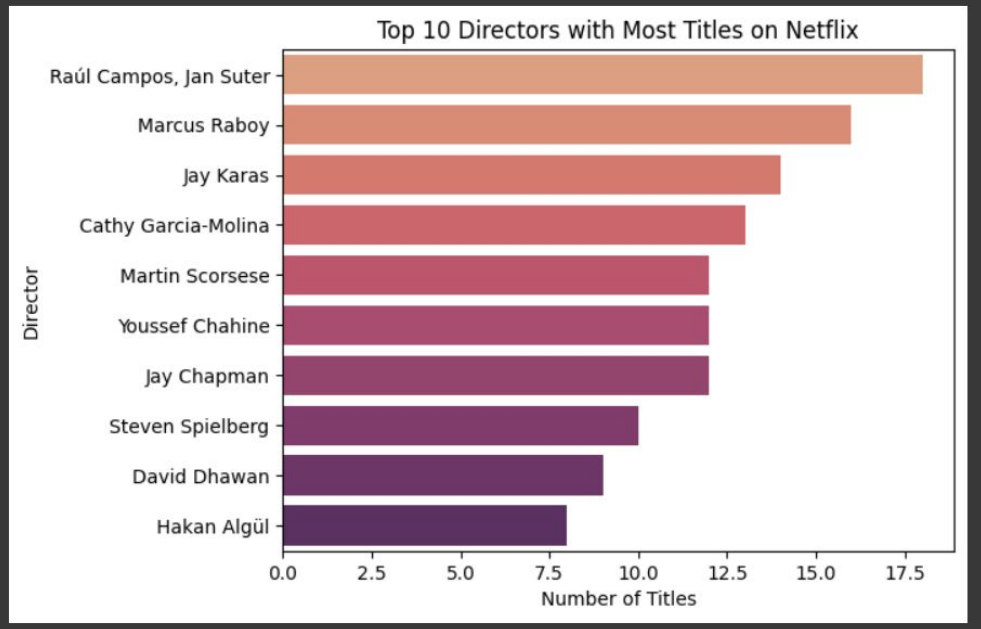
RESULTS2

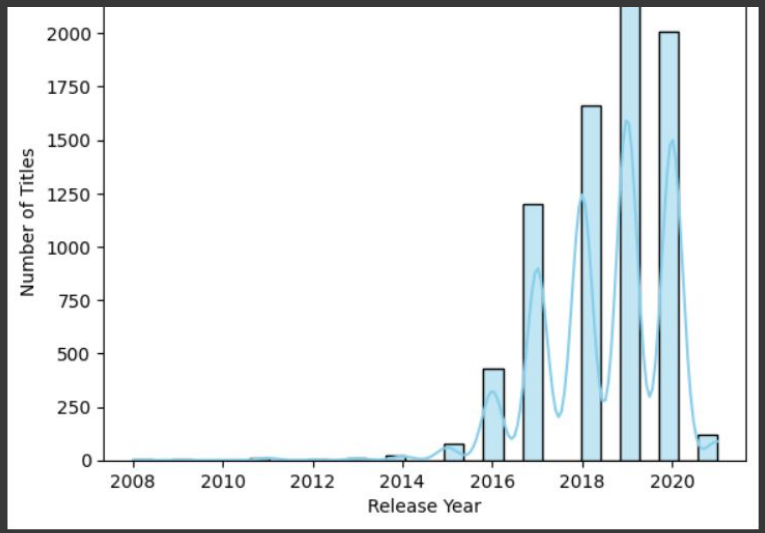
```
dtype: int64

🔹 Duplicate Rows: 2
```

```
[4]
#Descriptive Statistics
print("\n🔹 Descriptive Statistics (Numerical Columns):\n")
print(df.describe())

print("\n🔹 Descriptive Statistics (Categorical Columns):\n")
print(df.describe(include=['object']))
```

```
🔹 Descriptive Statistics (Numerical Columns):

       Show_Id Category              Title                Director  \
count     7789     7789               7789                    5401
unique    7787        2               7787                    4050
top       s6621    Movie  The Lost Okoroshi  Raúl Campos, Jan Suter
freq         2     5379                  2                      18

                      Cast        Country     Release_Date Rating  Duration  \
count                 7071           7282             7779   7782      7789
unique                6831            681             1565     14       216
top     David Attenborough  United States  January 1, 2020  TV-MA  1 Season
freq                    18           2556              118   2865      1608

                Type                                        Description
```

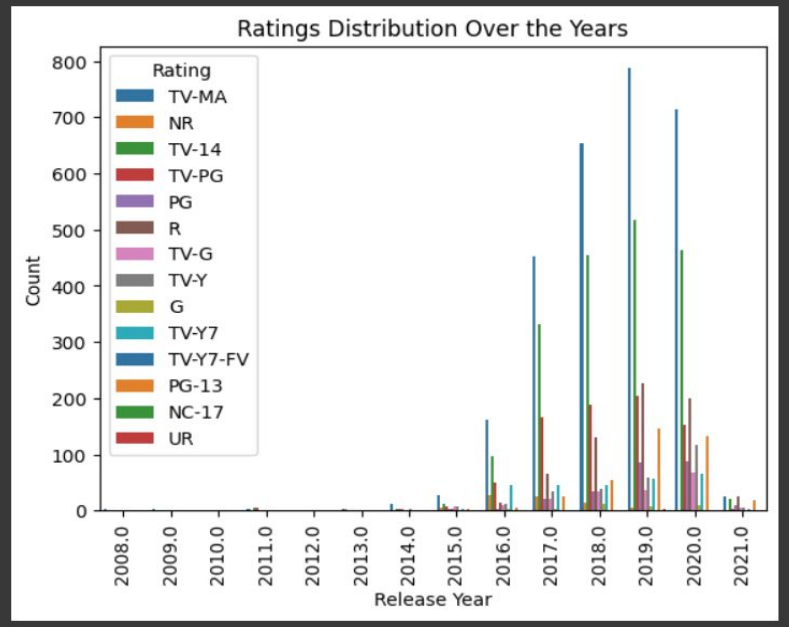
```
[26]
✓ 0s
top_directors = df['Director'].value_counts().head(10)
sns.barplot(x=top_directors.values, y=top_directors.index, palette='flare')
plt.title("Top 10 Directors with Most Titles on Netflix")
plt.xlabel("Number of Titles")
plt.ylabel("Director")
plt.show()
```





RESULTS3

```
plt.show()
```




```
#Key Insights Summary
print("\n KEY INSIGHTS:\n")

if 'type' in df.columns:
    movies = df[df['type'] == 'Movie'].shape[0]
    shows = df[df['type'] == 'TV Show'].shape[0]
    print(f"• Movies: {movies} | TV Shows: {shows}")

if 'country' in df.columns:
    top_country = df['country'].value_counts().idxmax()
    print(f"• Country with most titles: {top_country}")

if 'release_year' in df.columns:
    most_year = df['release_year'].value_counts().idxmax()
    print(f"• Peak year of content release: {most_year}")

if 'listed_in' in df.columns:
    top_genre = all_genres.value_counts().idxmax()
    print(f"• Most common genre: {top_genre}")
```