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

```
# Load Netflix dataset
df = pd.read_csv("netflix_titles.csv")
# Basic info
print(df.info())
# First few rows
print(df.head())
# Description of object and date columns
print(df.describe(include="all"))
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       8807 non-null   object
 1   type          8807 non-null   object
 2   title         8807 non-null   object
 3   director      6173 non-null   object
 4   cast          7982 non-null   object
 5   country       7976 non-null   object
 6   date_added    8797 non-null   object
 7   release_year  8807 non-null   int64
 8   rating        8803 non-null   object
 9   duration      8804 non-null   object
 10  listed_in     8807 non-null   object
 11  description   8807 non-null   object
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
None
  show_id     type                  title         director  \
0      s1    Movie   Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show           Blood & Water              NaN
2      s3  TV Show               Ganglands  Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans              NaN
4      s5  TV Show           Kota Factory              NaN

                                          cast        country  \
0                                          NaN  United States
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...   South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...            NaN
3                                          NaN            NaN
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...          India

          date_added  release_year rating   duration  \
0  September 25, 2021          2020  PG-13     90 min
1  September 24, 2021          2021  TV-MA  2 Seasons
2  September 24, 2021          2021  TV-MA   1 Season
3  September 24, 2021          2021  TV-MA   1 Season
4  September 24, 2021          2021  TV-MA  2 Seasons

                                     listed_in  \
0                                Documentaries
1   International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act...
3                            Docuseries, Reality TV
```

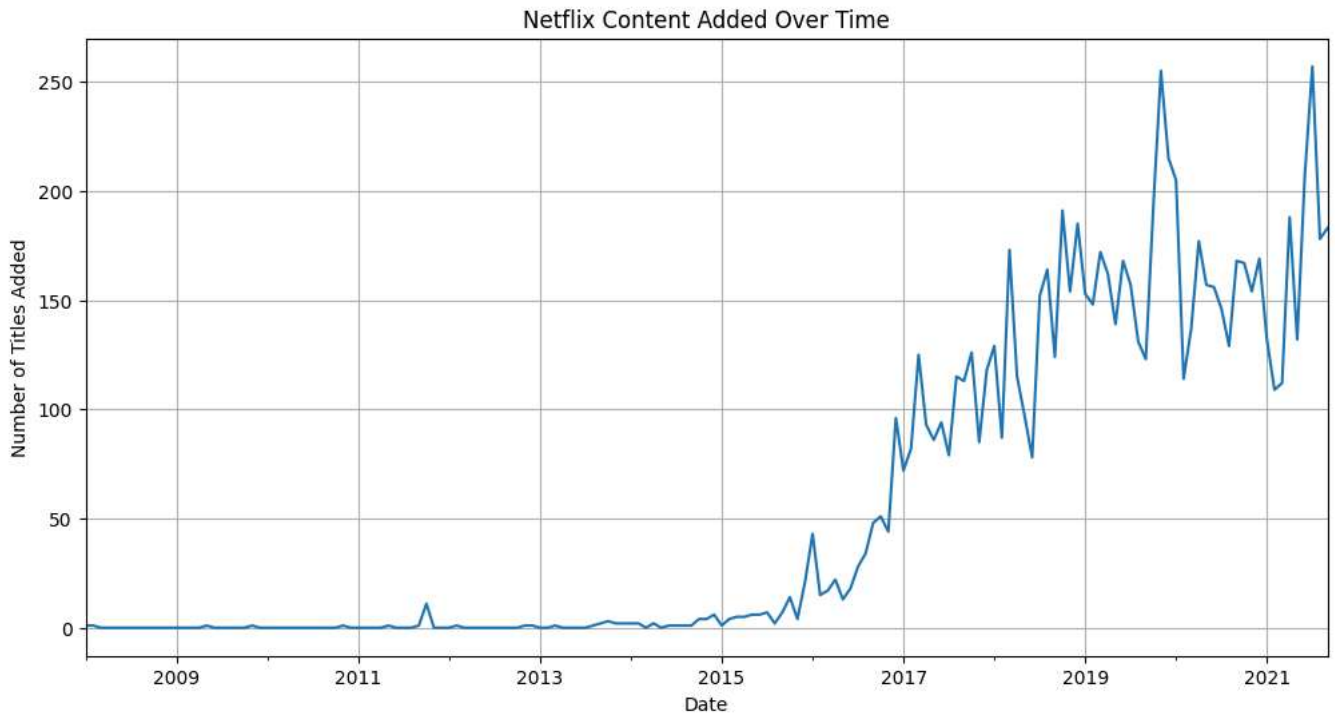
```
4  International TV Shows, Romantic TV Shows, TV ...

                                      description
0  As her father nears the end of his life, filmm...
1  After crossing paths at a party, a Cape Town t...
2  To protect his family from a powerful drug lor...
3  Feuds, flirtations and toilet talk go down amo...
4  In a city of coaching centers known to train I...
       show_id   type   title       director               cast  \
count     8807   8807    8807           6173               7982
unique    8807      2    8807           4528               7692
top      s8807  Movie  Zubaan  Rajiv Chilaka  David Attenborough
```

```
# Unique values
print("Number of unique countries:", df["country"].nunique())
print("Number of unique directors:", df["director"].nunique())
# Content counts by type (Movie/TV Show)
print(df["type"].value_counts())
# Count of content per year
print(df["release_year"].value_counts().head())
# Grouped by country and type
print(df.groupby(["country", "type"]).size().sort_values(ascending=False).head(10))
# Convert date_added to datetime
df["date_added"] = pd.to_datetime(df["date_added"], format="mixed", errors="coerce")
# Set index to date_added
df.set_index("date_added", inplace=True)
# Monthly content additions
monthly_content = df.resample("M").size()
```

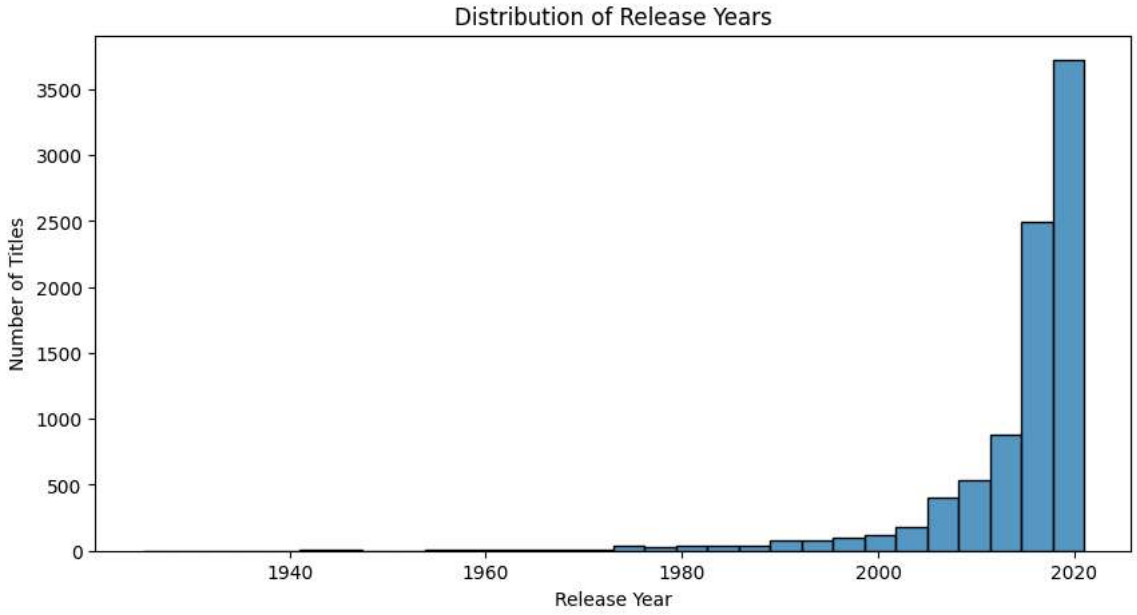
```
Number of unique countries: 748
Number of unique directors: 4528
type
Movie      6131
TV Show    2676
Name: count, dtype: int64
release_year
2018    1147
2017    1032
2019    1030
2020     953
2016     902
Name: count, dtype: int64
country         type
United States   Movie      2058
India           Movie       893
United States   TV Show     760
United Kingdom  TV Show     213
                Movie       206
Japan           TV Show     169
South Korea     TV Show     158
Canada          Movie       122
Spain           Movie        97
Egypt           Movie        92
dtype: int64
/tmp/ipython-input-1638353343.py:15: FutureWarning: 'M' is deprecated and will be removed in a future version, please use 'ME' instead.
  monthly_content = df.resample("M").size()
```

```
# Plotting
plt.figure(figsize=(12,6))
monthly_content.plot()
plt.title("Netflix Content Added Over Time")
```

```
plt.xlabel("Date")
plt.ylabel("Number of Titles Added")
plt.grid(True)
plt.show()
```

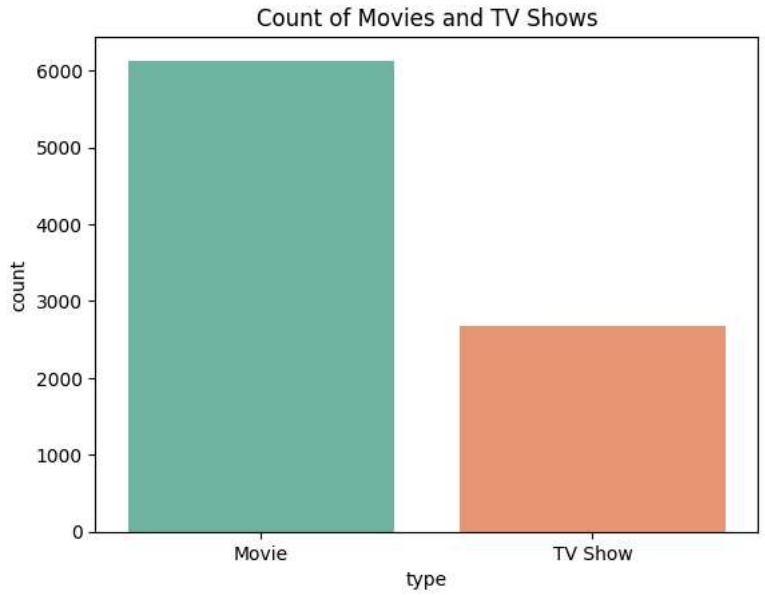



```
# Distribution of release years
plt.figure(figsize=(10,5))
sns.histplot(df["release_year"], bins=30, kde=False)
plt.title("Distribution of Release Years")
plt.xlabel("Release Year")
plt.ylabel("Number of Titles")
plt.show()
```

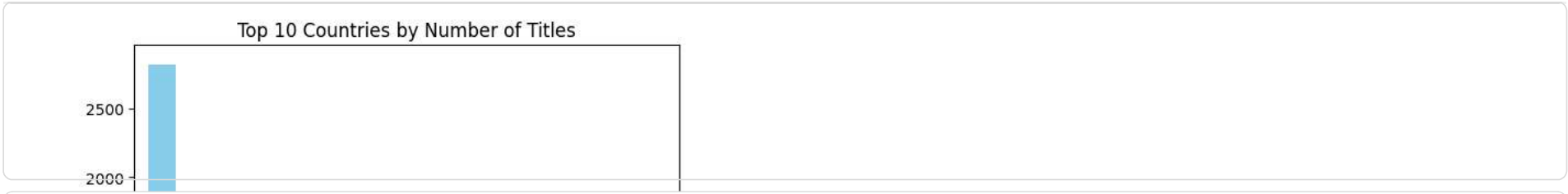



```
# Content Type Count
sns.countplot(data=df, x="type", palette="Set2")
plt.title("Count of Movies and TV Shows")
plt.show()
```

```
/tmp/ipython-input-1908447077.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 `legend=False` for the same effect.

  sns.countplot(data=df, x="type", palette="Set2")
```




```
# Top 10 countries with most content
top_countries = df["country"].value_counts().head(10)
top_countries.plot(kind="bar", color="skyblue")
plt.title("Top 10 Countries by Number of Titles")
plt.ylabel("Count")
plt.xticks(rotation=45)
plt.show()
```




```
# Genre Frequency
genres = df["listed_in"].str.split(", ", expand=True).stack()
top_genres = genres.value_counts().head(10)
top_genres.plot(kind="barh", color="coral")
plt.title("Top 10 Genres on Netflix")
plt.xlabel("Count")
plt.gca().invert_yaxis()
plt.show()
```