```
# This Python 3 environment comes with many helpful analytics
libraries installed
# It is defined by the kaggle/python Docker image:
https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/"
directory
# For example, running this (by clicking run or pressing Shift+Enter)
will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/)
that gets preserved as output when you create a version using "Save &
Run All"
# You can also write temporary files to /kaggle/temp/, but they won't
be saved outside of the current session
```

```
/kaggle/input/netflix-shows/netflix_titles.csv
```

Netflix Movies & TV Shows – Exploratory Data Analysis

Objective

The goal of this analysis is to explore Netflix's content catalog and identify trends in content type, release year, country distribution, and genres.

```
import os

for root, dirs, files in os.walk("/kaggle/input"):
    for name in files:
        print(os.path.join(root, name))
```

```
/kaggle/input/netflix-shows/netflix_titles.csv
```

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

```
sns.set(style="whitegrid")

df = pd.read_csv("/kaggle/input/netflix-shows/netflix_titles.csv")
df.head()
```

```
  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...
```

```
df.shape
df.info()
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
```

	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
freq	1	6131	1	19	19
mean	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN	NaN

	country	date_added	release_year	rating	duration
count	7976	8797	8807.000000	8803	8804
unique	748	1767	NaN	17	220
top	United States	January 1, 2020	NaN	TV-MA	1 Season
freq	2818	109	NaN	3207	1793
mean	NaN	NaN	2014.180198	NaN	NaN
std	NaN	NaN	8.819312	NaN	NaN
min	NaN	NaN	1925.000000	NaN	NaN
25%	NaN	NaN	2013.000000	NaN	NaN
50%	NaN	NaN	2017.000000	NaN	NaN
75%	NaN	NaN	2019.000000	NaN	NaN
max	NaN	NaN	2021.000000	NaN	NaN

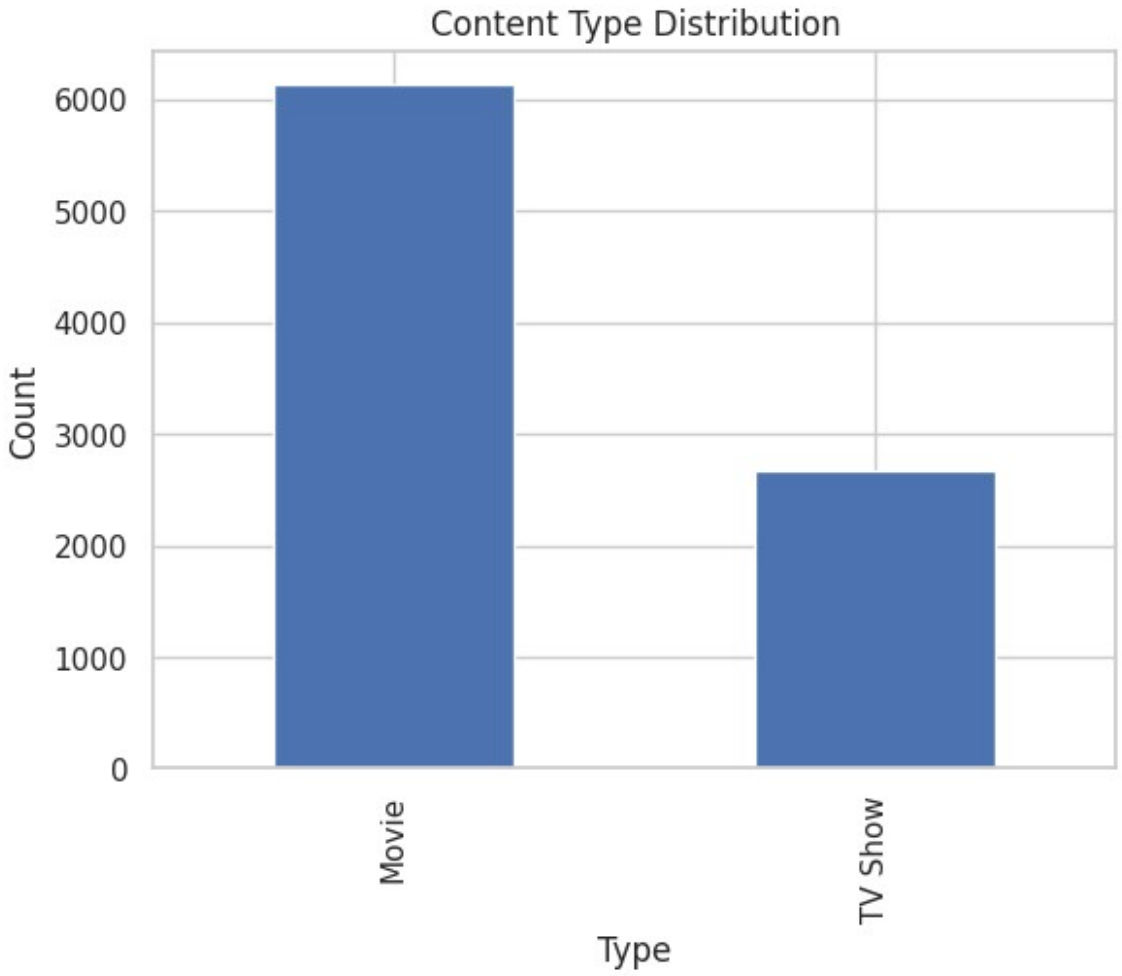
```
                              listed_in  \
count                              8807
unique                              514
top     Dramas, International Movies
freq                                362
mean                                NaN
std                                 NaN
min                                 NaN
25%                                 NaN
50%                                 NaN
75%                                 NaN
max                                 NaN

                                                 description
count                                                   8807
unique                                                  8775
top     Paranormal activity at a lush, abandoned prope...
freq                                                       4
mean                                                     NaN
std                                                      NaN
min                                                      NaN
25%                                                      NaN
50%                                                      NaN
75%                                                      NaN
max                                                      NaN
```

```
df.isnull().sum().sort_values(ascending=False)
```

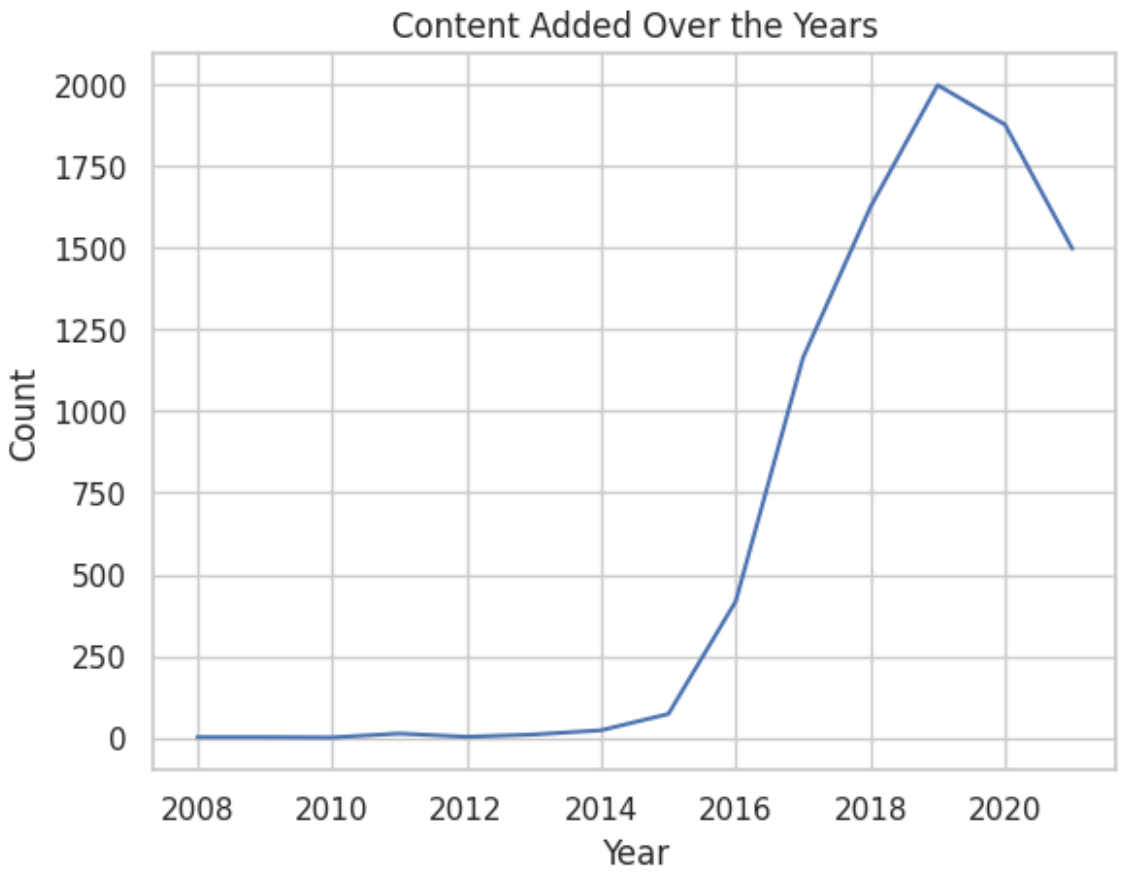
```
director        2634
country          831
cast             825
date_added        10
rating             4
duration           3
show_id            0
type               0
title              0
release_year       0
listed_in          0
description        0
dtype: int64
```

```
df['country'] = df['country'].fillna('Unknown')
df['rating'] = df['rating'].fillna('Not Rated')
df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')

df['type'].value_counts().plot(kind='bar')
plt.title('Content Type Distribution')
plt.xlabel('Type')
```

```
plt.ylabel('Count')
plt.show()
```




```
df['year_added'] = df['date_added'].dt.year

df['year_added'].value_counts().sort_index().plot(kind='line')
plt.title('Content Added Over the Years')
plt.xlabel('Year')
plt.ylabel('Count')
plt.show()
```




```
content_by_year = df['year_added'].value_counts().sort_index()
content_by_year.head()
```

```
year_added
2008.0     2
2009.0     2
2010.0     1
2011.0    13
2012.0     3
Name: count, dtype: int64
```

```
df['type'].value_counts()
```

```
type
Movie      6131
TV Show    2676
Name: count, dtype: int64
```

```
country_series = df['country'].dropna().str.split(', ').explode()
top_countries = country_series.value_counts().head(10)
top_countries
```

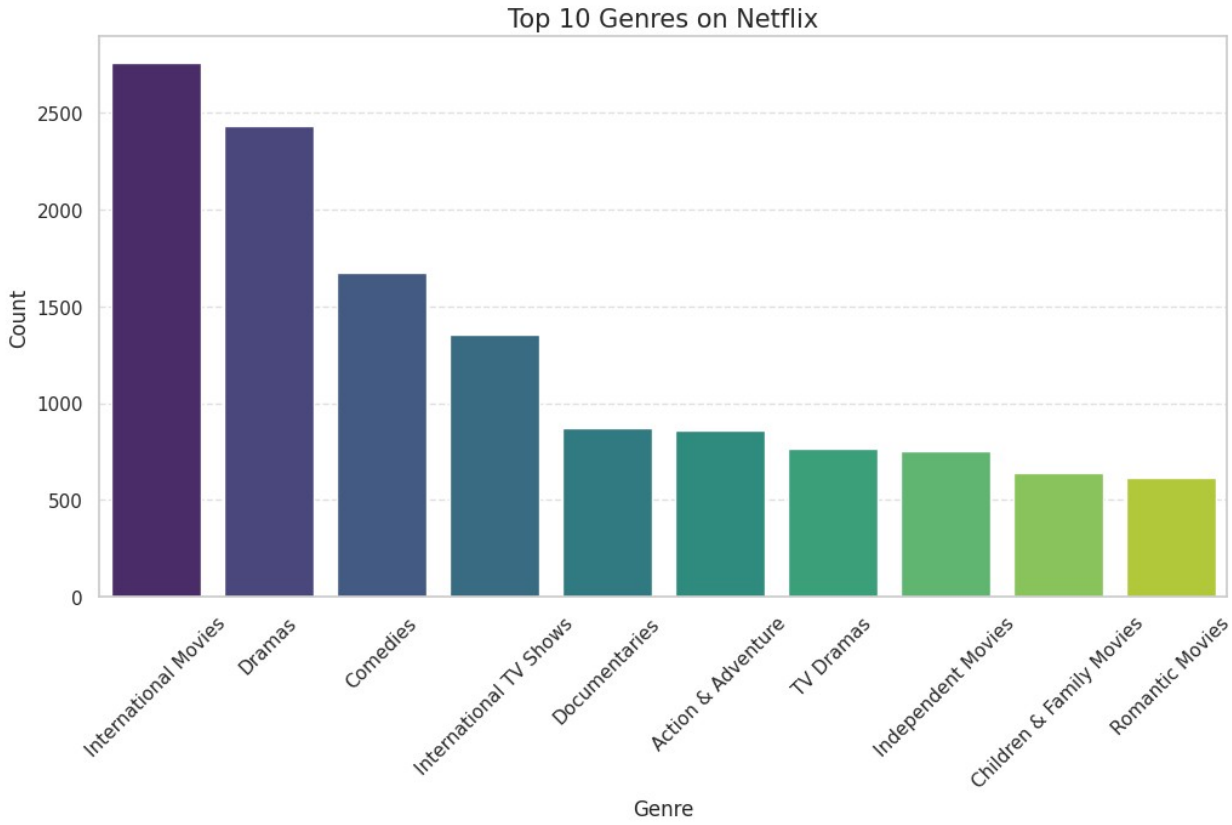
```
country
United States     3689
India             1046
Unknown            831
United Kingdom     804
Canada             445
France             393
Japan              318
Spain              232
South Korea        231
Germany            226
Name: count, dtype: int64
```

```
genre_series = df['listed_in'].str.split(', ').explode()
top_genres = genre_series.value_counts().head(10)
top_genres
```

```
listed_in
International Movies        2752
Dramas                      2427
Comedies                    1674
International TV Shows      1351
Documentaries                869
Action & Adventure           859
TV Dramas                    763
Independent Movies           756
Children & Family Movies     641
Romantic Movies              616
Name: count, dtype: int64
```

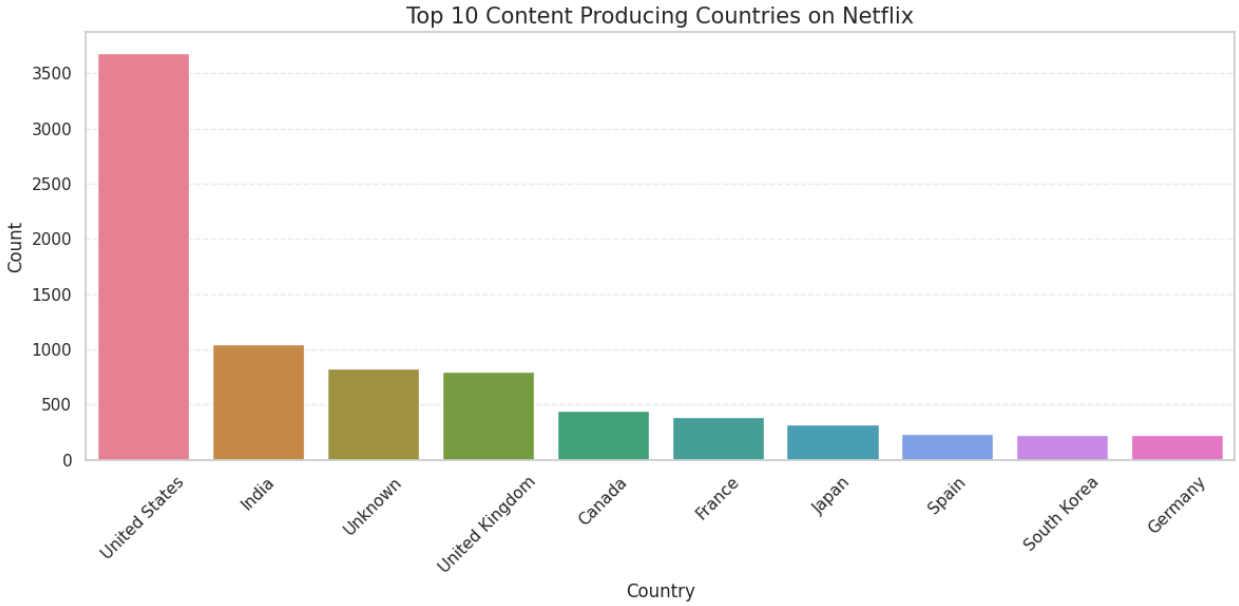
```
import matplotlib.pyplot as plt
import seaborn as sns

genres = df['listed_in'].str.split(',
').explode().value_counts().head(10)

plt.figure(figsize=(12, 6))

sns.barplot(
    x=genres.index,
    y=genres.values,
    palette='viridis',
    hue=genres.index,
    legend=False
)

plt.title('Top 10 Genres on Netflix', fontsize=15)
plt.xlabel('Genre', fontsize=12)
plt.ylabel('Count', fontsize=12)
plt.xticks(rotation=45)
```

```
plt.grid(axis='y', linestyle='--', alpha=0.6)
plt.show()
```




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

top_countries = df['country'].str.split(',
').explode().value_counts().head(10)

plt.figure(figsize=(12, 6))

sns.barplot(
    x=top_countries.index,
    y=top_countries.values,
    palette='husl',
    hue=top_countries.index,
    legend=False
)

plt.title('Top 10 Content Producing Countries on Netflix',
fontsize=15)
plt.xlabel('Country', fontsize=12)
plt.ylabel('Count', fontsize=12)
```

```
plt.xticks(rotation=45)

plt.grid(axis='y', linestyle='--', alpha=0.4)
plt.tight_layout()
plt.show()
```




```
print("Movies make up a larger portion of Netflix's catalog compared to TV Shows.")
```

```
Movies make up a larger portion of Netflix's catalog compared to TV Shows.
```