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

# Load the dataset
df = pd.read_csv(r'C:\\Users\\LENOVO\\Desktop\\ds assignment\\
netflix_titles.csv')

df.head(40)

   show_id     type                                              title
\
0       s1    Movie                               Dick Johnson Is Dead

1       s2  TV Show                                      Blood & Water

2       s3  TV Show                                          Ganglands

3       s4  TV Show                              Jailbirds New Orleans

4       s5  TV Show                                       Kota Factory

5       s6  TV Show                                      Midnight Mass

6       s7    Movie                   My Little Pony: A New Generation

7       s8    Movie                                            Sankofa

8       s9  TV Show                      The Great British Baking Show

9      s10    Movie                                       The Starling

10     s11  TV Show                Vendetta: Truth, Lies and The Mafia

11     s12  TV Show                                   Bangkok Breaking

12     s13    Movie                                       Je Suis Karl

13     s14    Movie                   Confessions of an Invisible Girl

14     s15  TV Show                    Crime Stories: India Detectives

15     s16  TV Show                                  Dear White People

16     s17    Movie  Europe's Most Dangerous Man: Otto Skorzeny in ...

17     s18  TV Show                                    Falsa identidad

18     s19    Movie                                          Intrusion

19     s20  TV Show                                             Jaguar
```

```
20     s21  TV Show    Monsters Inside: The 24 Faces of Billy Milligan
21     s22  TV Show                             Resurrection: Ertugrul
22     s23    Movie                                    Avvai Shanmughi
23     s24    Movie       Go! Go! Cory Carson: Chrissy Takes the Wheel
24     s25    Movie                                              Jeans
25     s26  TV Show                               Love on the Spectrum
26     s27    Movie                                     Minsara Kanavu
27     s28    Movie                                          Grown Ups
28     s29    Movie                                         Dark Skies
29     s30    Movie                                           Paranoia
30     s31    Movie                                    Ankahi Kahaniya
31     s32  TV Show                                 Chicago Party Aunt
32     s33  TV Show                                      Sex Education
33     s34  TV Show                                         Squid Game
34     s35  TV Show                            Tayo and Little Wizards
35     s36    Movie                     The Father Who Moves Mountains
36     s37    Movie                                     The Stronghold
37     s38  TV Show                                        Angry Birds
38     s39    Movie                                Birth of the Dragon
39     s40  TV Show                                       Chhota Bheem

                                             director  \
0                                     Kirsten Johnson
1                                                 NaN
2                                     Julien Leclercq
3                                                 NaN
4                                                 NaN
5                                       Mike Flanagan
6                       Robert Cullen, José Luis Ucha
7                                        Haile Gerima
```

```
8                                     Andy Devonshire
9                                      Theodore Melfi
10                                                NaN
11                                  Kongkiat Komesiri
12                                Christian Schwochow
13                                      Bruno Garotti
14                                                NaN
15                                                NaN
16      Pedro de Echave García, Pablo Azorín Williams
17                                                NaN
18                                         Adam Salky
19                                                NaN
20                                    Olivier Megaton
21                                                NaN
22                                     K.S. Ravikumar
23                            Alex Woo, Stanley Moore
24                                         S. Shankar
25                                                NaN
26                                        Rajiv Menon
27                                       Dennis Dugan
28                                      Scott Stewart
29                                     Robert Luketic
30  Ashwiny Iyer Tiwari, Abhishek Chaubey, Saket C...
31                                                NaN
32                                                NaN
33                                                NaN
34                                                NaN
35                                       Daniel Sandu
36                                     Cédric Jimenez
37                                                NaN
38                                       George Nolfi
39                                                NaN

                                                 cast  \
0                                                 NaN
1   Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...
2   Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...
3                                                 NaN
4   Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...
5   Kate Siegel, Zach Gilford, Hamish Linklater, H...
6   Vanessa Hudgens, Kimiko Glenn, James Marsden, ...
7   Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
8   Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...
9   Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...
10                                                NaN
11  Sukollawat Kanarot, Sushar Manaying, Pavarit M...
12  Luna Wedler, Jannis Niewöhner, Milan Peschel, ...
13  Klara Castanho, Lucca Picon, Júlia Gomes, Marc...
14                                                NaN
```

```
15  Logan Browning, Brandon P. Bell, DeRon Horton,...
16                                                NaN
17  Luis Ernesto Franco, Camila Sodi, Sergio Goyri...
18  Freida Pinto, Logan Marshall-Green, Robert Joh...
19  Blanca Suárez, Iván Marcos, Óscar Casas, Adriá...
20                                                NaN
21  Engin Altan Düzyatan, Serdar Gökhan, Hülya Dar...
22  Kamal Hassan, Meena, Gemini Ganesan, Heera Raj...
23  Maisie Benson, Paul Killam, Kerry Gudjohnsen, ...
24  Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...
25                                   Brooke Satchwell
26  Arvind Swamy, Kajol, Prabhu Deva, Nassar, S.P....
27  Adam Sandler, Kevin James, Chris Rock, David S...
28  Keri Russell, Josh Hamilton, J.K. Simmons, Dak...
29  Liam Hemsworth, Gary Oldman, Amber Heard, Harr...
30  Abhishek Banerjee, Rinku Rajguru, Delzad Hiwal...
31  Lauren Ash, Rory O'Malley, RuPaul Charles, Jil...
32  Asa Butterfield, Gillian Anderson, Ncuti Gatwa...
33  Lee Jung-jae, Park Hae-soo, Wi Ha-jun, Oh Youn...
34  Dami Lee, Jason Lee, Bommie Catherine Han, Jen...
35  Adrian Titieni, Elena Purea, Judith State, Val...
36  Gilles Lellouche, Karim Leklou, François Civil...
37  Antti Pääkkönen, Heljä Heikkinen, Lynne Guagli...
38  Billy Magnussen, Ron Yuan, Qu Jingjing, Terry ...
39  Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...

                                              country
date_added  \
0                                       United States  September 25,
2021
1                                        South Africa  September 24,
2021
2                                                 NaN  September 24,
2021
3                                                 NaN  September 24,
2021
4                                               India  September 24,
2021
5                                                 NaN  September 24,
2021
6                                                 NaN  September 24,
2021
7   United States, Ghana, Burkina Faso, United Kin...  September 24,
2021
8                                      United Kingdom  September 24,
2021
9                                       United States  September 24,
2021
10                                                NaN  September 24,
2021
11                                                NaN  September 23,
2021
12                            Germany, Czech Republic  September 23,
2021
13                                                NaN  September 22,
2021
14                                                NaN  September 22,
2021
15                                      United States  September 22,
2021
16                                                NaN  September 22,
2021
17                                             Mexico  September 22,
2021
18                                                NaN  September 22,
2021
19                                                NaN  September 22,
2021
20                                                NaN  September 22,
2021
21                                             Turkey  September 22,
2021
22                                                NaN  September 21,
2021
23                                                NaN  September 21,
2021
24                                              India  September 21,
2021
25                                          Australia  September 21,
2021
26                                                NaN  September 21,
2021
27                                      United States  September 20,
2021
28                                      United States  September 19,
2021
29                       United States, India, France  September 19,
2021
30                                                NaN  September 17,
2021
31                                                NaN  September 17,
2021
32                                     United Kingdom  September 17,
2021
33                                                NaN  September 17,
2021
34                                                NaN  September 17,
2021
```

```
35                                                NaN  September 17,
2021
36                                                NaN  September 17,
2021
37                                            Finland  September 16,
2021
38                       China, Canada, United States  September 16,
2021
39                                              India  September 16,
2021

    release_year rating   duration  \
0           2020  PG-13     90 min
1           2021  TV-MA  2 Seasons
2           2021  TV-MA   1 Season
3           2021  TV-MA   1 Season
4           2021  TV-MA  2 Seasons
5           2021  TV-MA   1 Season
6           2021     PG     91 min
7           1993  TV-MA    125 min
8           2021  TV-14  9 Seasons
9           2021  PG-13    104 min
10          2021  TV-MA   1 Season
11          2021  TV-MA   1 Season
12          2021  TV-MA    127 min
13          2021  TV-PG     91 min
14          2021  TV-MA   1 Season
15          2021  TV-MA  4 Seasons
16          2020  TV-MA     67 min
17          2020  TV-MA  2 Seasons
18          2021  TV-14     94 min
19          2021  TV-MA   1 Season
20          2021  TV-14   1 Season
21          2018  TV-14  5 Seasons
22          1996  TV-PG    161 min
23          2021   TV-Y     61 min
24          1998  TV-14    166 min
25          2021  TV-14  2 Seasons
26          1997  TV-PG    147 min
27          2010  PG-13    103 min
28          2013  PG-13     97 min
29          2013  PG-13    106 min
30          2021  TV-14    111 min
31          2021  TV-MA   1 Season
32          2020  TV-MA  3 Seasons
33          2021  TV-MA   1 Season
34          2020  TV-Y7   1 Season
35          2021  TV-MA    110 min
36          2021  TV-MA    105 min
```

```
37          2018  TV-Y7   1 Season
38          2017  PG-13     96 min
39          2021  TV-Y7  3 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 ...   
5                   TV Dramas, TV Horror, TV Mysteries   
6                             Children & Family Movies   
7     Dramas, Independent Movies, International Movies   
8                         British TV Shows, Reality TV   
9                                     Comedies, Dramas   
10   Crime TV Shows, Docuseries, International TV S...   
11   Crime TV Shows, International TV Shows, TV Act...   
12                        Dramas, International Movies   
13                  Children & Family Movies, Comedies   
14        British TV Shows, Crime TV Shows, Docuseries   
15                              TV Comedies, TV Dramas   
16                 Documentaries, International Movies   
17   Crime TV Shows, Spanish-Language TV Shows, TV ...   
18                                           Thrillers   
19   International TV Shows, Spanish-Language TV Sh...   
20   Crime TV Shows, Docuseries, International TV S...   
21   International TV Shows, TV Action & Adventure,...   
22                      Comedies, International Movies   
23                            Children & Family Movies   
24     Comedies, International Movies, Romantic Movies   
25      Docuseries, International TV Shows, Reality TV   
26    Comedies, International Movies, Music & Musicals   
27                                            Comedies   
28                     Horror Movies, Sci-Fi & Fantasy   
29                                           Thrillers   
30    Dramas, Independent Movies, International Movies   
31                                         TV Comedies   
32   British TV Shows, International TV Shows, TV C...   
33     International TV Shows, TV Dramas, TV Thrillers   
34                                            Kids' TV   
35             Dramas, International Movies, Thrillers   
36    Action & Adventure, Dramas, International Movies   
37                               Kids' TV, TV Comedies   
38                          Action & Adventure, Dramas   
39                                            Kids' 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...
5    The arrival of a charismatic young priest brin...
6    Equestria's divided. But a bright-eyed hero be...
7    On a photo shoot in Ghana, an American model s...
8    A talented batch of amateur bakers face off in...
9    A woman adjusting to life after a loss contend...
10   Sicily boasts a bold "Anti-Mafia" coalition. B...
11   Struggling to earn a living in Bangkok, a man ...
12   After most of her family is murdered in a terr...
13   When the clever but socially-awkward Tetê join...
14   Cameras following Bengaluru police on the job ...
15   Students of color navigate the daily slights a...
16   Declassified documents reveal the post-WWII li...
17   Strangers Diego and Isabel flee their home in ...
18   After a deadly home invasion at a couple’s new...
19   In the 1960s, a Holocaust survivor joins a gro...
20   In the late 1970s, an accused serial rapist cl...
21   When a good deed unwittingly endangers his cla...
22   Newly divorced and denied visitation rights wi...
23   From arcade games to sled days and hiccup cure...
24   When the father of the man she loves insists t...
25   Finding love can be hard for anyone. For young...
26   A tangled love triangle ensues when a man fall...
27   Mourning the loss of their beloved junior high...
28   A family’s idyllic suburban life shatters when...
29   Blackmailed by his company's CEO, a low-level ...
30   As big city life buzzes around them, lonely so...
31   Chicago Party Aunt Diane is an idolized troubl...
32   Insecure Otis has all the answers when it come...
33   Hundreds of cash-strapped players accept a str...
34   Tayo speeds into an adventure when his friends...
35   When his son goes missing during a snowy hike ...
36   Tired of the small-time grind, three Marseille...
37   Birds Red, Chuck and their feathered friends h...
38   A young Bruce Lee angers kung fu traditionalis...
39   A brave, energetic little boy with superhuman ...
```

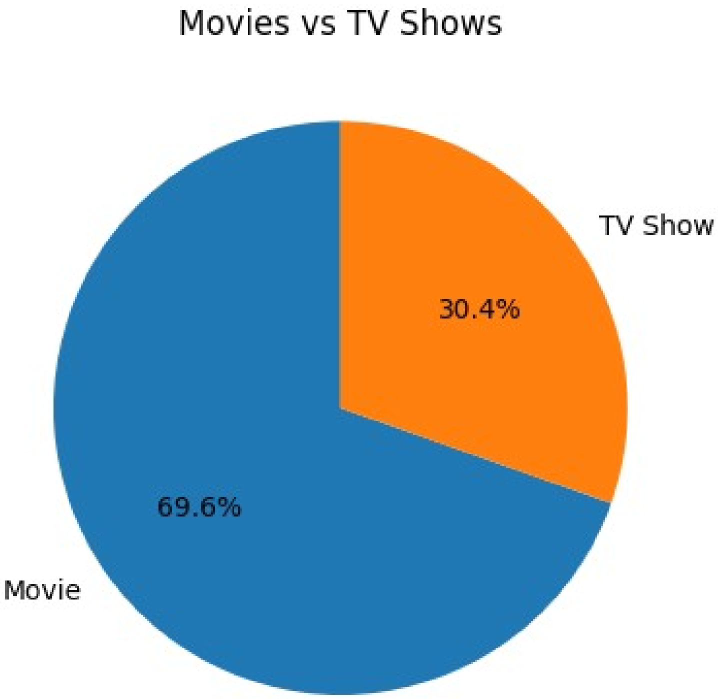
```
df.info()
```

```
<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
```

content Stratergy

```
# 1. Ratio of Movies vs TV Shows
type_counts = df['type'].value_counts()
print("Movies vs TV Shows:")
print(type_counts)

# Plot
type_counts.plot(kind='pie', autopct='%1.1f%%', startangle=90,
title='Movies vs TV Shows')
plt.ylabel('')
plt.show()
```

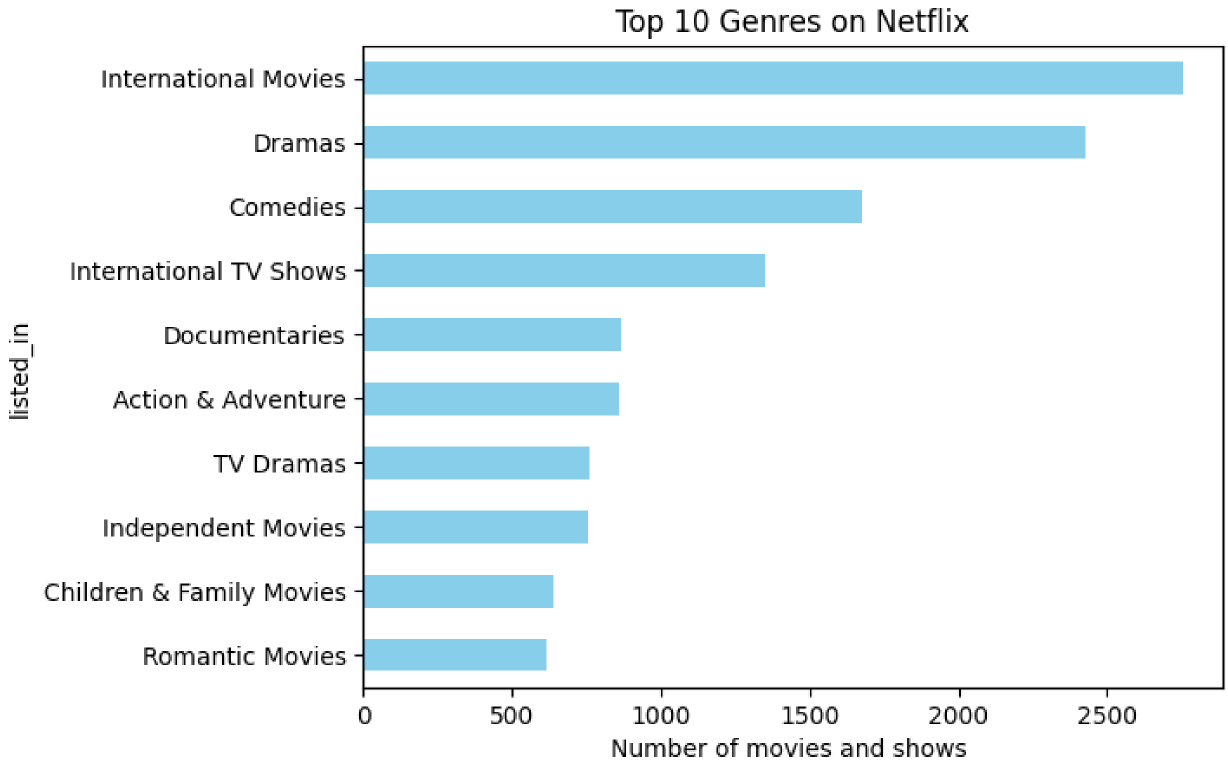
```
Movies vs TV Shows:
type
Movie      6131
TV Show    2676
Name: count, dtype: int64
```



```
# 2. Most popular genres on Netflix
# Genres are listed as comma-separated values; we need to split them
genre_series = df['listed_in'].dropna().str.split(', ')
genres = genre_series.explode().value_counts().head(10)
print("\nTop 10 Genres:")
print(genres)

# Plot
genres.plot(kind='barh', title='Top 10 Genres on Netflix',
color='skyblue')
plt.xlabel('Number of movies and shows ')
plt.gca().invert_yaxis()
plt.show()
```

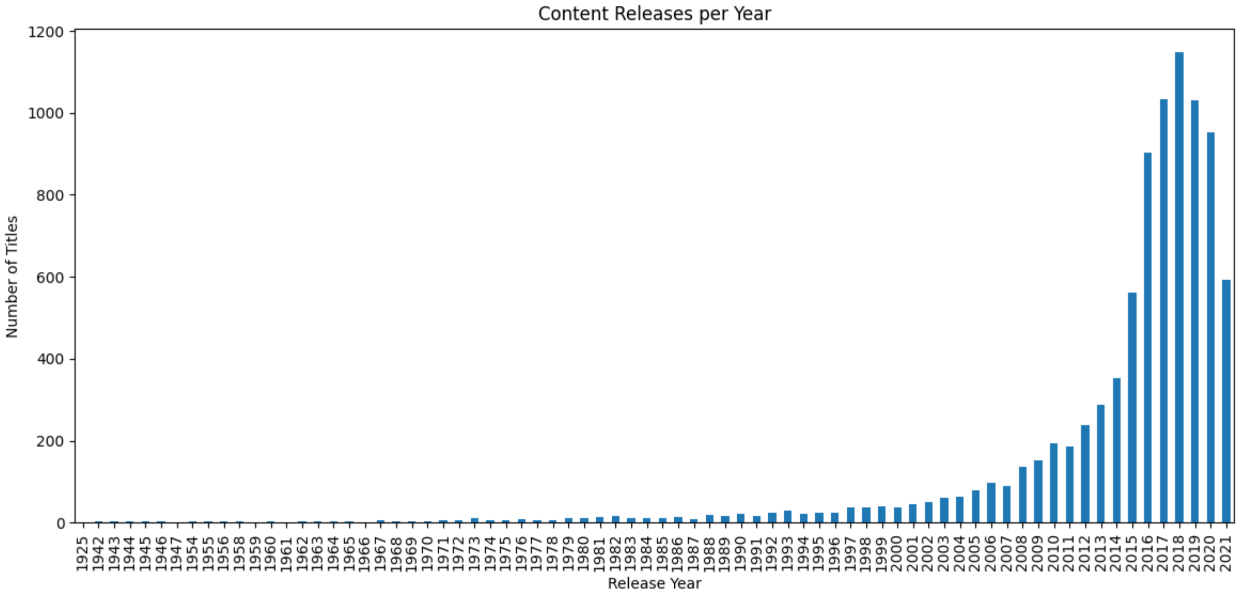
```
Top 10 Genres:
listed_in
International Movies        2752
Dramas                      2427
Comedies                    1674
International TV Shows      1351
Documentaries                869
Action & Adventure           859
TV Dramas                    763
Independent Movies           756
Children & Family Movies     641
```

```
Romantic Movies                 616
Name: count, dtype: int64
```



```
# 3. Years with the highest release of content
release_counts = df['release_year'].value_counts().sort_index()
print("\nContent Release by Year:")
print(release_counts.tail(10))

# Plot
release_counts.plot(kind='bar', figsize=(14, 6), title='Content
Releases per Year')
plt.xlabel('Release Year')
plt.ylabel('Number of Titles')
plt.show()


Content Release by Year:
release_year
2012     237
2013     288
2014     352
2015     560
2016     902
2017    1032
2018    1147
```

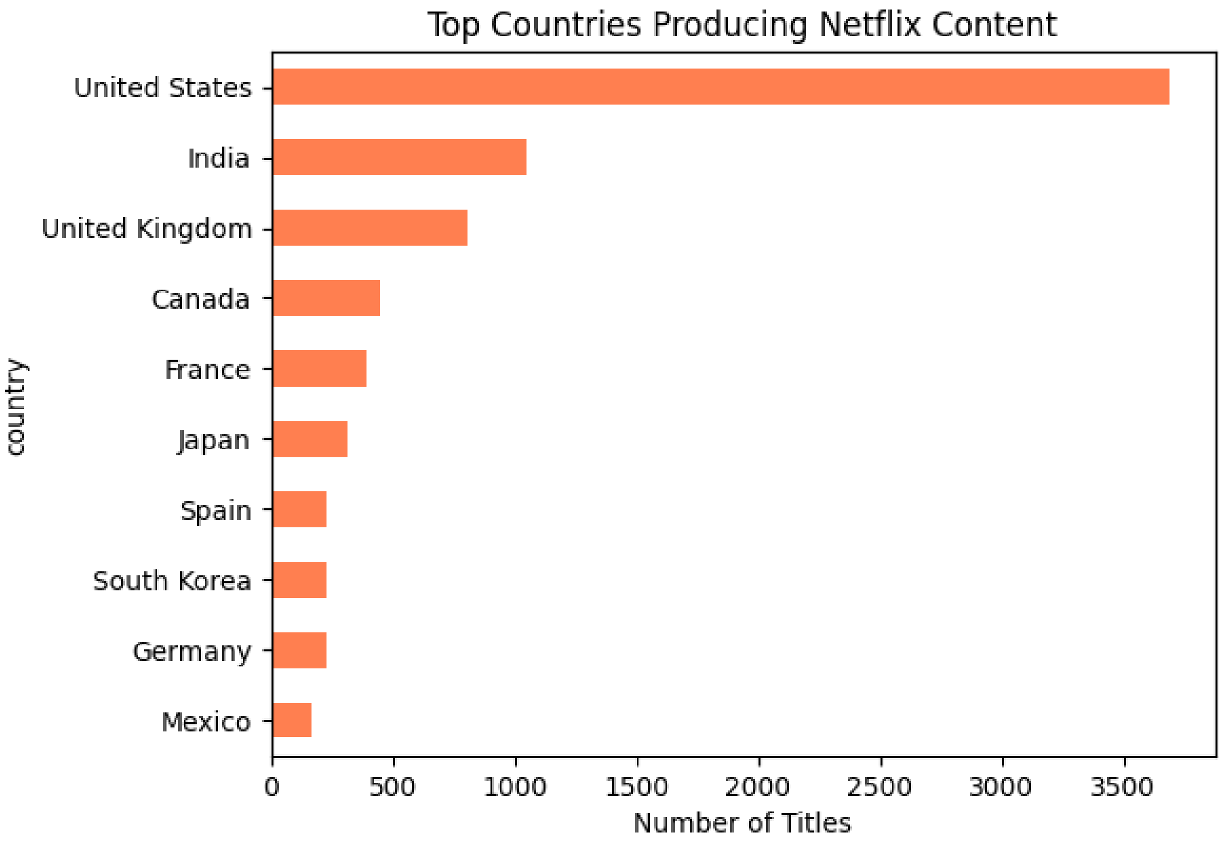
```
2019    1030
2020     953
2021     592
Name: count, dtype: int64
```



```
# 4. Countries producing the most content
country_series = df['country'].dropna().str.split(', ')
countries = country_series.explode().value_counts().head(10)
print("\nTop Content-Producing Countries:")
print(countries)

# Plot
countries.plot(kind='barh', title='Top Countries Producing Netflix Content', color='coral')
plt.xlabel('Number of movies and shows')
plt.gca().invert_yaxis()
plt.show()
```

```
Top Content-Producing Countries:
country
United States     3689
India             1046
United Kingdom     804
Canada             445
France             393
Japan              318
Spain              232
South Korea        231
Germany            226
```

```
Mexico               169
Name: count, dtype: int64
```



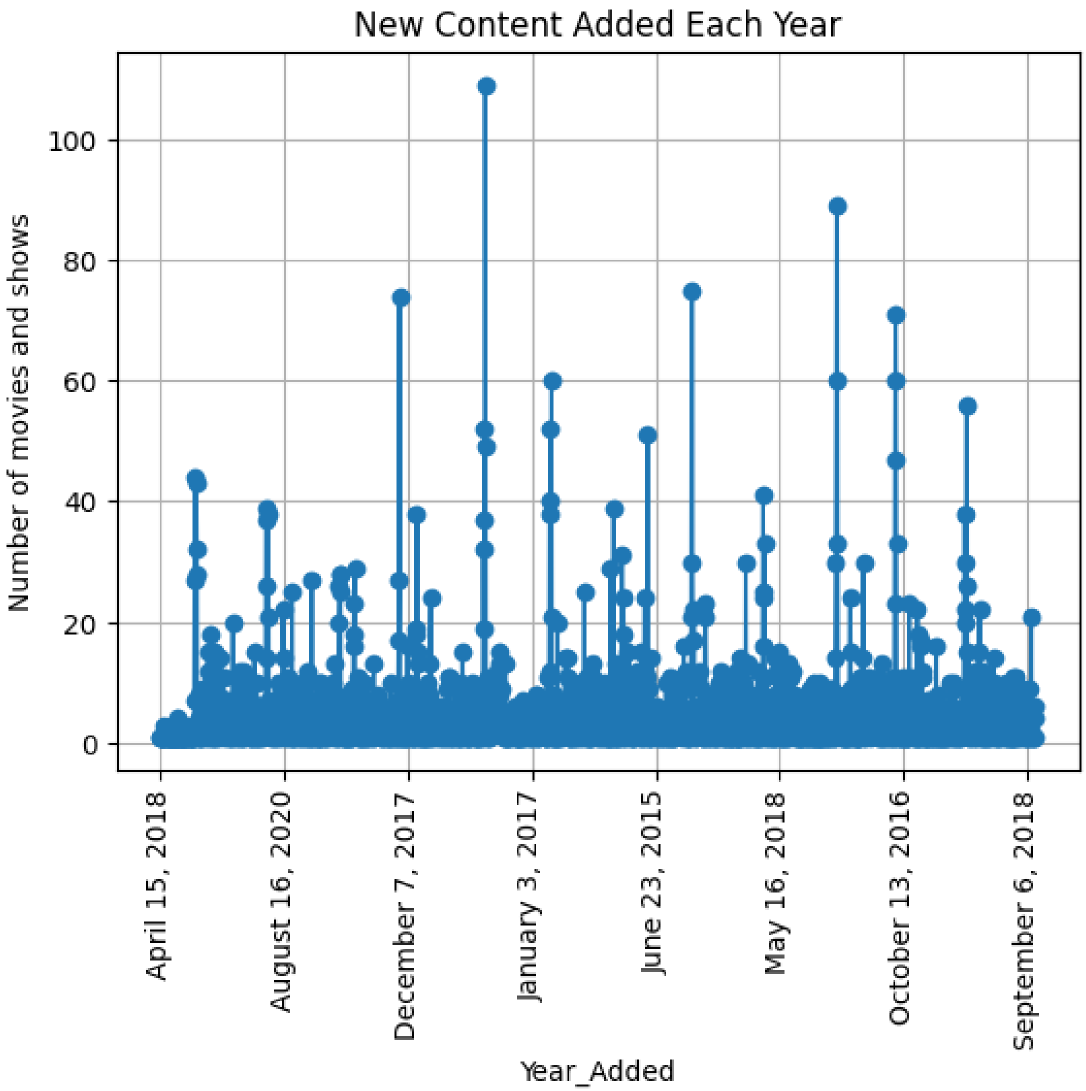
```
df.head()

  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
year_added
0  As her father nears the end of his life, filmm...  September 25,
2021
1  After crossing paths at a party, a Cape Town t...  September 24,
2021
2  To protect his family from a powerful drug lor...  September 24,
2021
3  Feuds, flirtations and toilet talk go down amo...  September 24,
2021
4  In a city of coaching centers known to train I...  September 24,
2021
```

```
# 5. Trend of adding content year by year
df['year_added'] = df['date_added']
added_per_year = df['year_added'].value_counts().sort_index()
print("\nTrend of Content Added Each Year:")
print(added_per_year)

# Plot
added_per_year.plot(kind='line', marker='o', title='New Content Added
Each Year')
plt.xlabel('Year_Added')
plt.ylabel('Number of movies and shows')
plt.xticks(rotation=90)
plt.grid(True)
plt.show()
```

```
Trend of Content Added Each Year:
year_added
 April 15, 2018      1
 April 16, 2019      1
 April 17, 2016      1
 April 20, 2017      1
 April 4, 2017       1
```

```
                          ..
September 9, 2016     1
September 9, 2018     1
September 9, 2019     1
September 9, 2020     6
September 9, 2021     4
Name: count, Length: 1767, dtype: int64
```

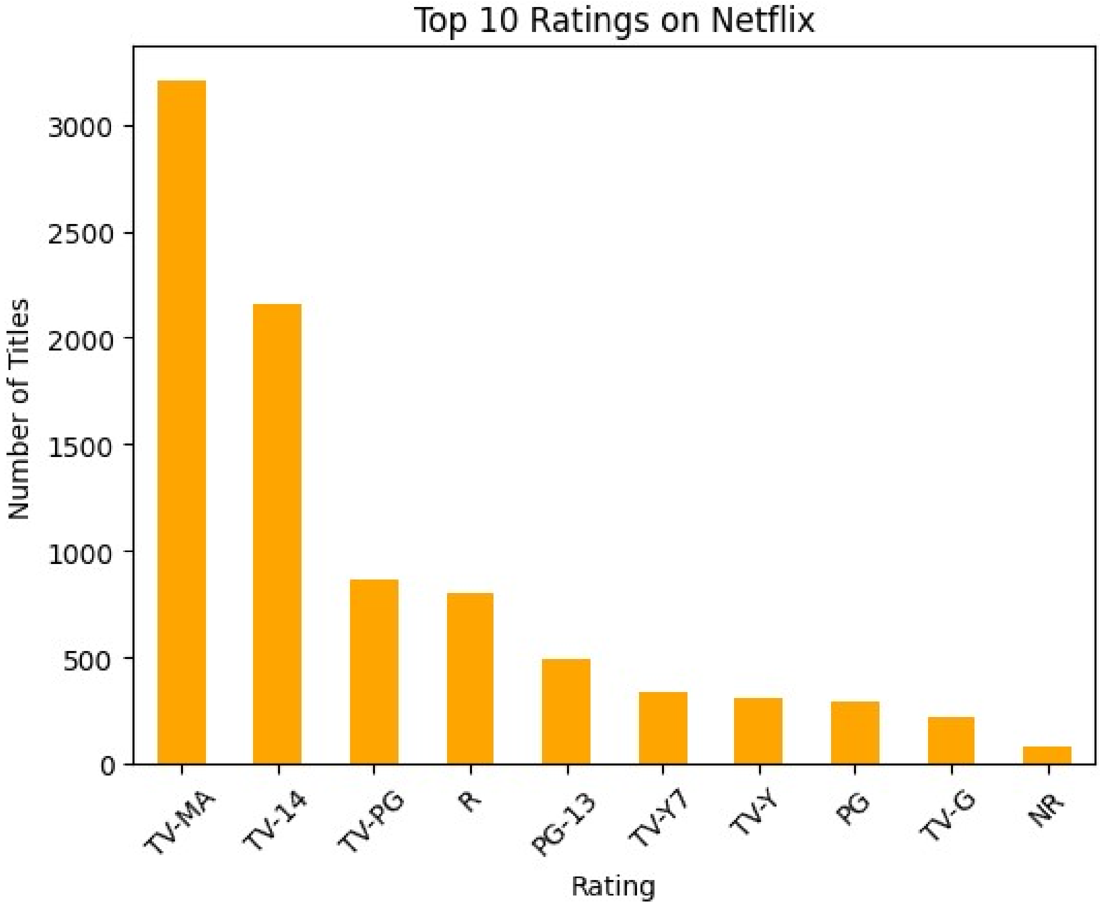


Use Demographics & Targeting

```
# 6. Most frequent ratings on Netflix
rating_counts = df['rating'].value_counts().head(10)
print("Most Frequent Ratings on Netflix:")
print(rating_counts)
```

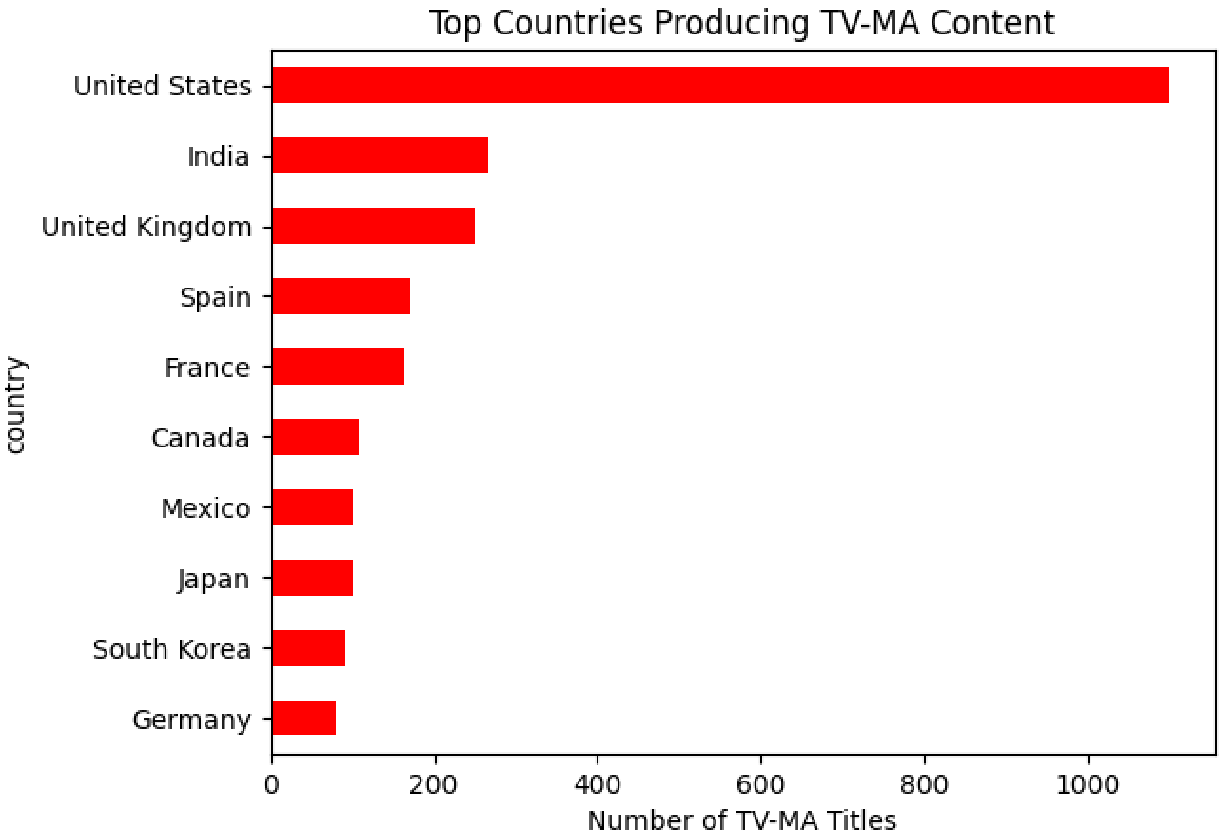
```
# Plot
rating_counts.plot(kind='bar', color='orange', title='Top 10 Ratings
on Netflix')
plt.xlabel('Rating')
plt.ylabel('Number of Titles')
plt.xticks(rotation=45)
plt.show()
```

```
Most Frequent Ratings on Netflix:
rating
TV-MA    3207
TV-14    2160
TV-PG     863
R         799
PG-13     490
TV-Y7     334
TV-Y      307
PG        287
TV-G      220
NR         80
Name: count, dtype: int64
```



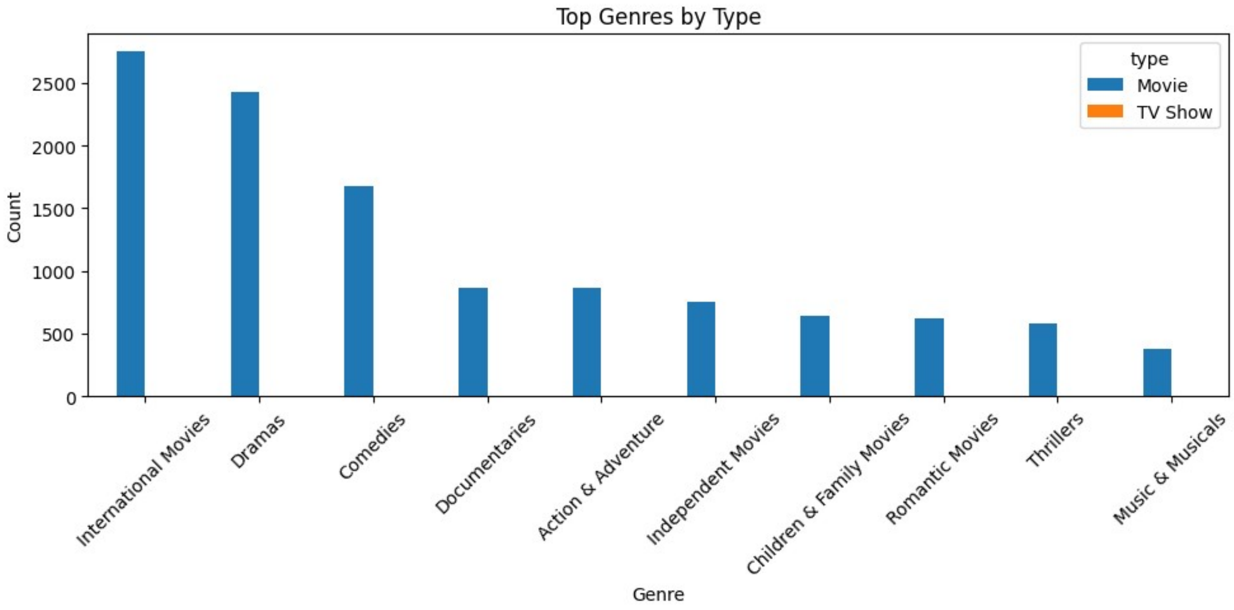
```
# 7. Countries that produce more mature content (TV-MA)
tv_ma_df = df[df['rating'] == 'TV-MA']
tv_ma_countries = tv_ma_df['country'].dropna().str.split(',
').explode().value_counts().head(10)
print("\nTop Countries Producing TV-MA Content:")
print(tv_ma_countries)

# Plot
tv_ma_countries.plot(kind='barh', color='red', title='Top Countries
Producing TV-MA Content')
plt.xlabel('Number of TV-MA Titles')
plt.gca().invert_yaxis()
plt.show()
```

```
Top Countries Producing TV-MA Content:
country
United States     1100
India              266
United Kingdom     251
Spain              170
France             163
Canada             107
Mexico             102
Japan              101
South Korea         92
Germany             79
Name: count, dtype: int64
```



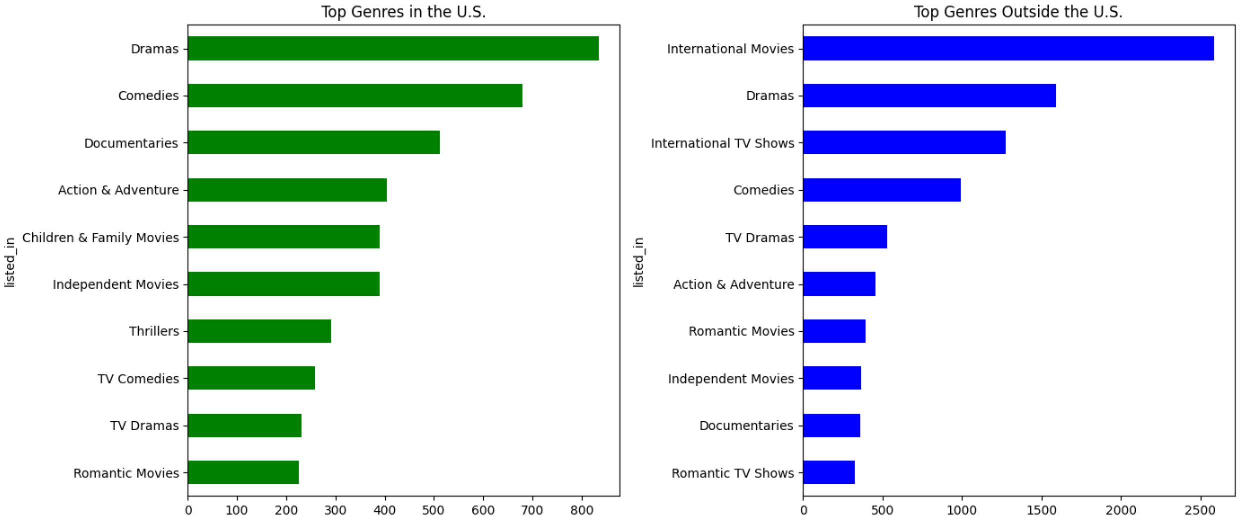
```python
# 8. Genres associated with TV Shows vs Movies
genre_split = df[['type', 'listed_in']].dropna()
genre_split =
genre_split.assign(listed_in=genre_split['listed_in'].str.split(', '))
genre_exploded = genre_split.explode('listed_in')
genre_pivot = genre_exploded.groupby(['listed_in',
'type']).size().unstack().fillna(0)
genre_pivot = genre_pivot.sort_values(by='Movie' if 'Movie' in
genre_pivot else 0, ascending=False)

print("\nGenre Distribution Across Types:")
print(genre_pivot.head(10))

# Plot
genre_pivot.head(10).plot(kind='bar', figsize=(10, 5), title='Top
Genres by Type')
plt.xlabel('Genre')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

```
Genre Distribution Across Types:
type                        Movie  TV Show
listed_in
International Movies       2752.0      0.0
Dramas                     2427.0      0.0
Comedies                   1674.0      0.0
Documentaries               869.0      0.0
Action & Adventure          859.0      0.0
Independent Movies          756.0      0.0
Children & Family Movies    641.0      0.0
Romantic Movies             616.0      0.0
Thrillers                   577.0      0.0
Music & Musicals            375.0      0.0
```



```python
# 9. Dominant genres in the U.S. vs Other Countries
df['is_US'] = df['country'].fillna('').apply(lambda x: 'United States' in x)
us_genres = df[df['is_US']]['listed_in'].dropna().str.split(', ').explode().value_counts().head(10)
non_us_genres = df[~df['is_US']]['listed_in'].dropna().str.split(', ').explode().value_counts().head(10)

# Plot
fig, axs = plt.subplots(1, 2, figsize=(14, 6))
us_genres.plot(kind='barh', ax=axs[0], title='Top Genres in the U.S.', color='green')
non_us_genres.plot(kind='barh', ax=axs[1], title='Top Genres Outside the U.S.', color='blue')
axs[0].invert_yaxis()
```

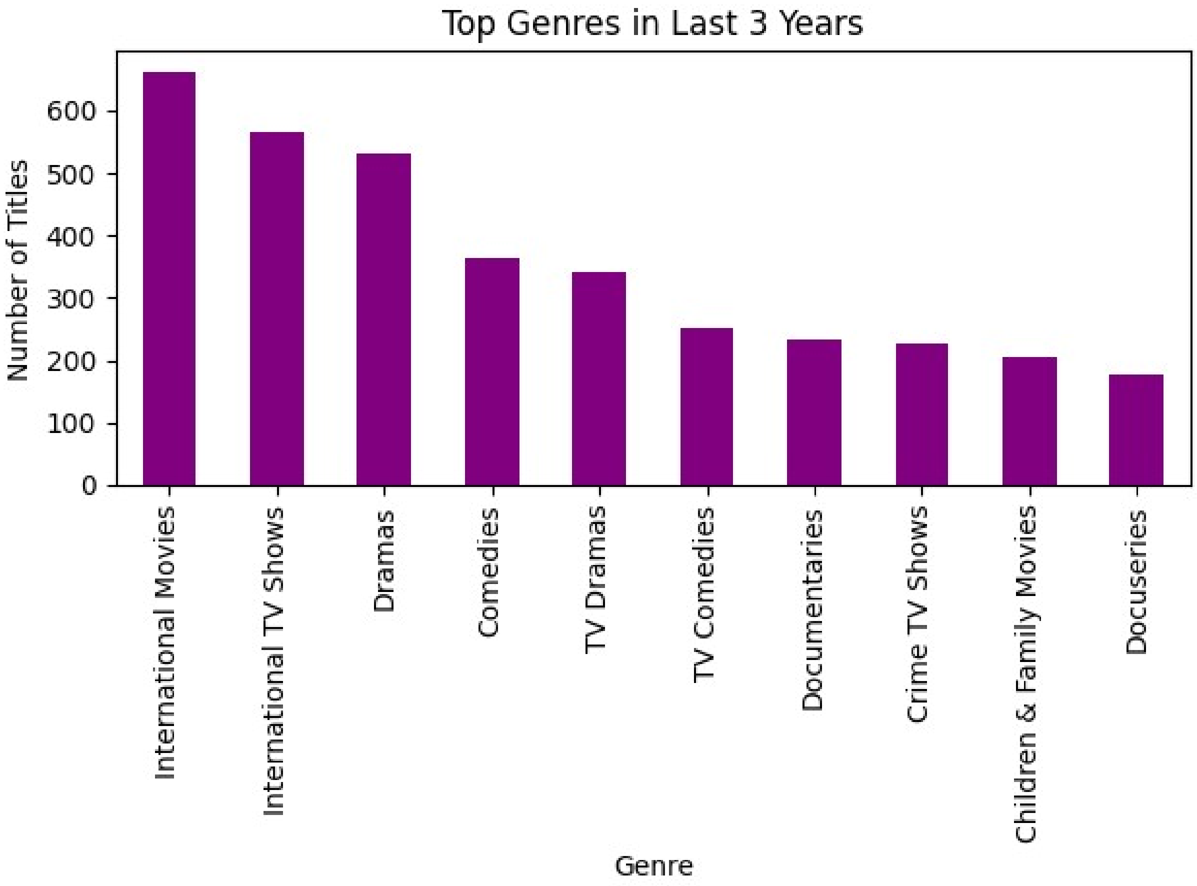
```
axs[1].invert_yaxis()
plt.tight_layout()
plt.show()
```



```
# 10. Most popular genres in the last 3 years
latest_year = df['release_year'].max()
recent_df = df[df['release_year'] >= latest_year - 2]
recent_genres = recent_df['listed_in'].dropna().str.split(',
').explode().value_counts().head(10)

print("\nTop Genres in Last 3 Years:")
print(recent_genres)

# Plot
recent_genres.plot(kind='bar', color='purple', title='Top Genres in
Last 3 Years')
plt.xlabel('Genre')
plt.ylabel('Number of Titles')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```

```
Top Genres in Last 3 Years:
listed_in
International Movies      662
International TV Shows    564
Dramas                    530
Comedies                  362
TV Dramas                 343
TV Comedies               252
Documentaries             234
```

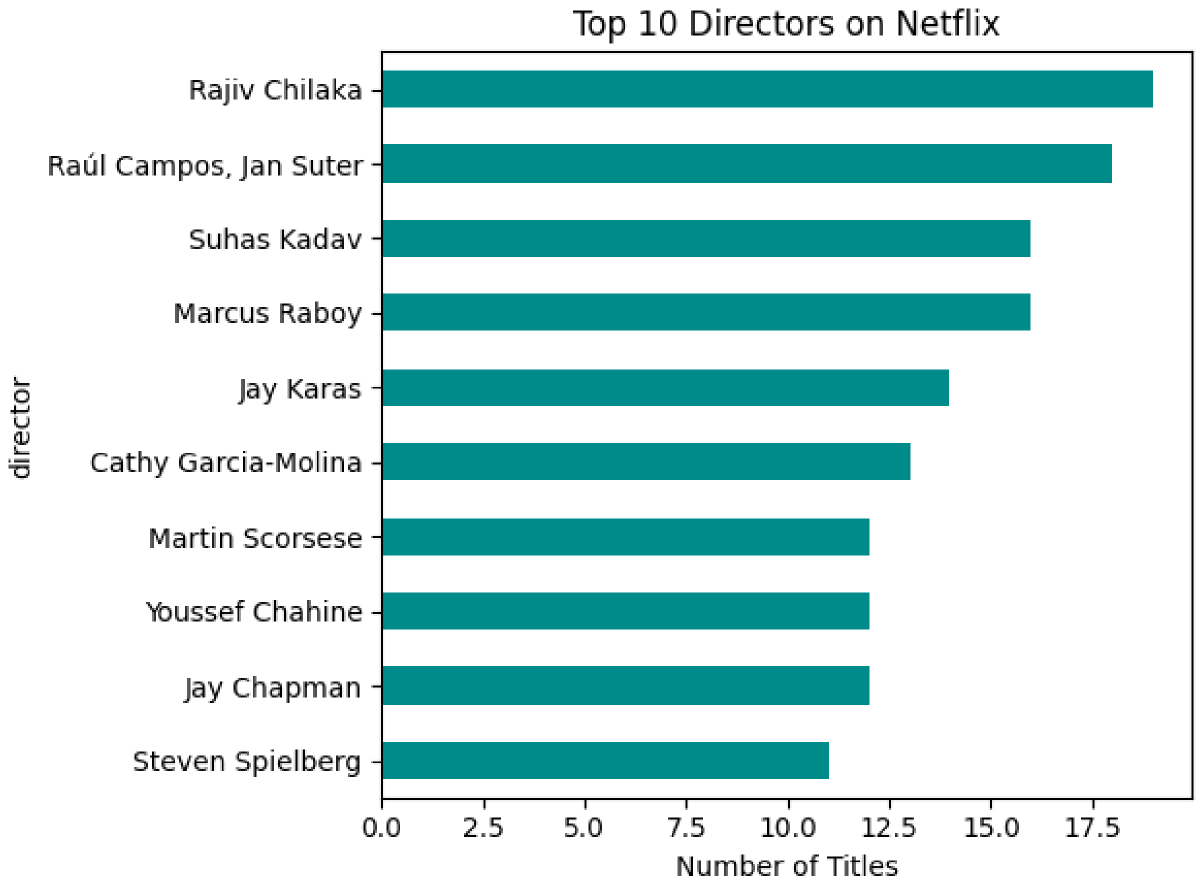
```
Crime TV Shows              226
Children & Family Movies    205
Docuseries                  178
Name: count, dtype: int64
```



Talent Acquisition & Partnership

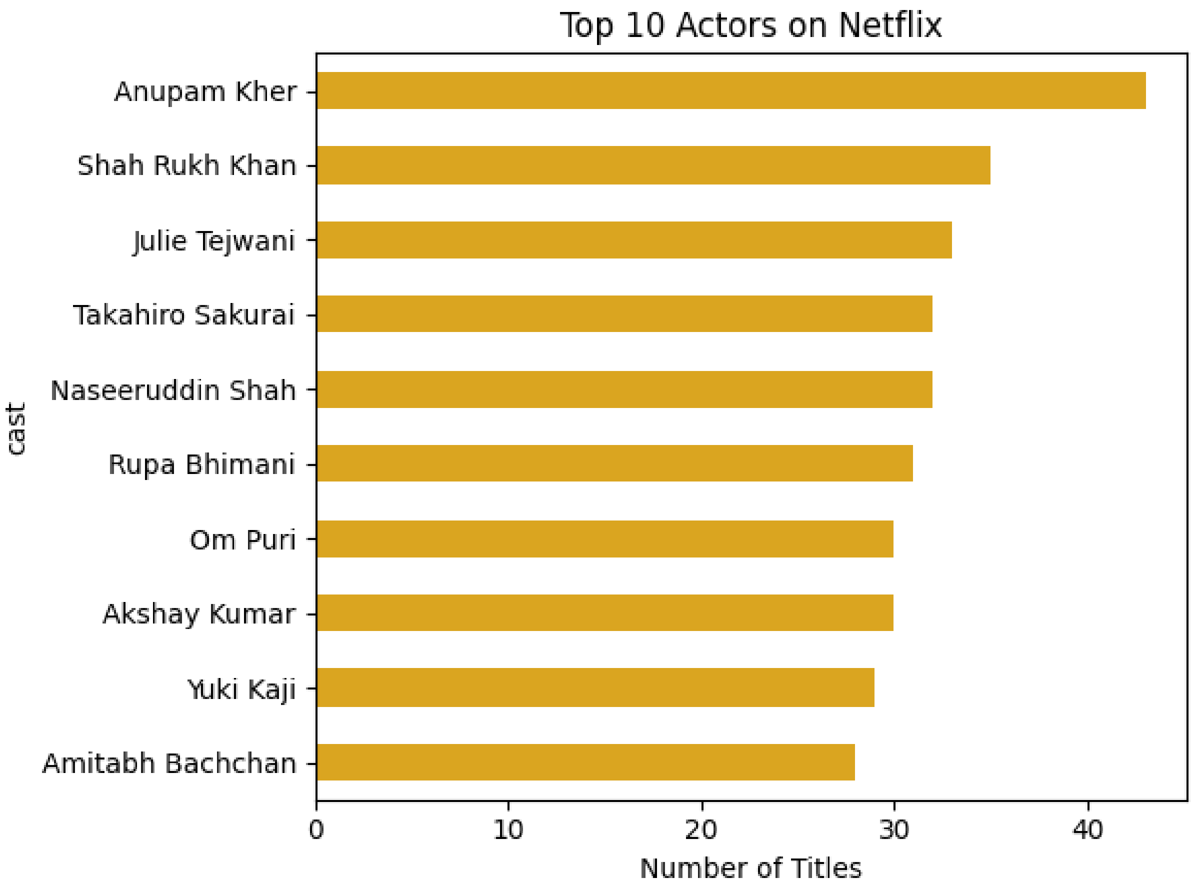
```
# 11. Top 10 Directors with Most Netflix Content
top_directors = df['director'].dropna().value_counts().head(10)
print("Top 10 Directors on Netflix:")
print(top_directors)

# Plot
top_directors.plot(kind='barh', title='Top 10 Directors on Netflix',
color='darkcyan')
plt.xlabel('Number of Titles')
plt.gca().invert_yaxis()
plt.tight_layout()
plt.show()
```

```
Top 10 Directors on Netflix:
director
Rajiv Chilaka                19
Raúl Campos, Jan Suter       18
Suhas Kadav                  16
Marcus Raboy                 16
Jay Karas                    14
Cathy Garcia-Molina          13
Martin Scorsese              12
Youssef Chahine              12
Jay Chapman                  12
Steven Spielberg             11
Name: count, dtype: int64
```



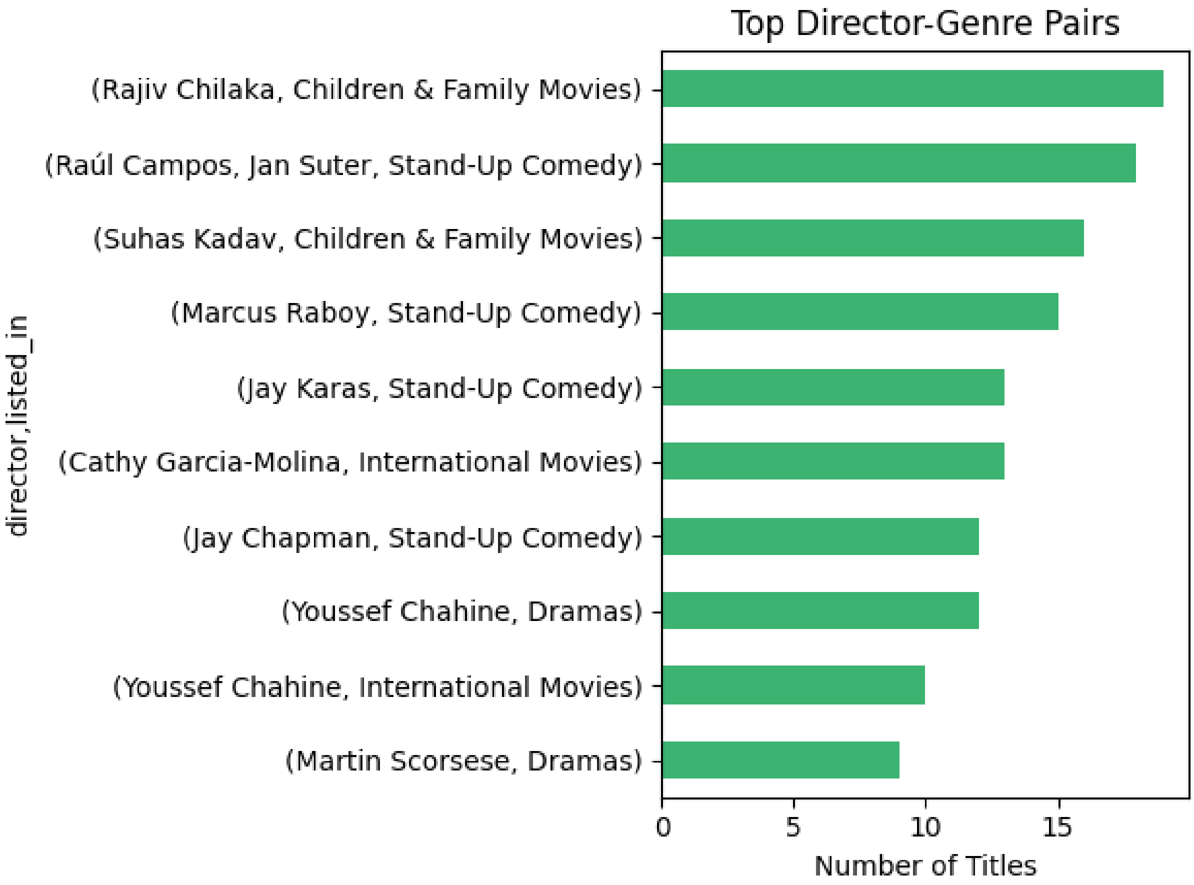
```
# 12. Most Frequent Actors in Netflix Shows
actor_series = df['cast'].dropna().str.split(', ')
actors = actor_series.explode().str.strip()
top_actors = actors.value_counts().head(10)
print("\nTop 10 Actors on Netflix:")
print(top_actors)
```

```
# Plot
top_actors.plot(kind='barh', title='Top 10 Actors on Netflix',
color='goldenrod')
plt.xlabel('Number of Titles')
plt.gca().invert_yaxis()
plt.tight_layout()
plt.show()


Top 10 Actors on Netflix:
cast
Anupam Kher         43
Shah Rukh Khan      35
Julie Tejwani       33
Takahiro Sakurai    32
Naseeruddin Shah    32
Rupa Bhimani        31
Om Puri             30
Akshay Kumar        30
Yuki Kaji           29
Amitabh Bachchan    28
Name: count, dtype: int64
```



```
# 13. Most Frequent Director-Genre Pairs
dir_genre = df[['director', 'listed_in']].dropna()
dir_genre =
dir_genre.assign(listed_in=dir_genre['listed_in'].str.split(', '))
dir_genre_expanded = dir_genre.explode('listed_in')
pair_counts = dir_genre_expanded.groupby(['director',
'listed_in']).size().sort_values(ascending=False).head(10)
print("\nTop Director-Genre Pairs:")
print(pair_counts)

# Plot
pair_counts.plot(kind='barh', title='Top Director-Genre Pairs',
color='mediumseagreen')
plt.xlabel('Number of Titles')
plt.gca().invert_yaxis()
plt.tight_layout()
plt.show()
```

```
Top Director-Genre Pairs:
director                 listed_in
Rajiv Chilaka            Children & Family Movies    19
Raúl Campos, Jan Suter  Stand-Up Comedy             18
Suhas Kadav              Children & Family Movies    16
Marcus Raboy             Stand-Up Comedy             15
Jay Karas                Stand-Up Comedy             13
Cathy Garcia-Molina      International Movies        13
Jay Chapman              Stand-Up Comedy             12
Youssef Chahine          Dramas                      12
                         International Movies        10
Martin Scorsese          Dramas                       9
dtype: int64
```



```
# 14. Titles with Unknown Directors or Cast
unknown_directors = df['director'].isna().sum()
unknown_cast = df['cast'].isna().sum()
print(f"\nNumber of Titles with Unknown Directors:
{unknown_directors}")
print(f"Number of Titles with Unknown Cast Members: {unknown_cast}")
```

```
Number of Titles with Unknown Directors: 2634
Number of Titles with Unknown Cast Members: 825
```

duration & Engagement

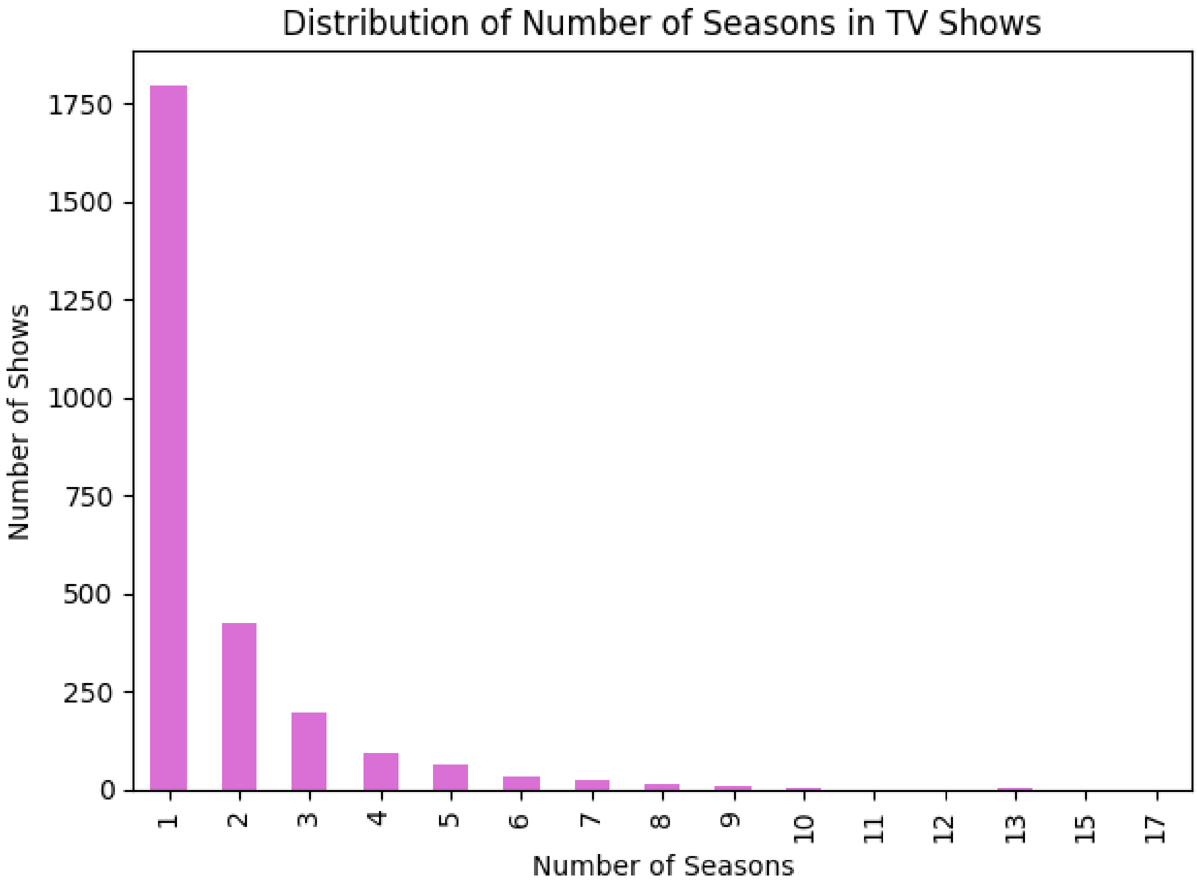
```
# 15. Average duration of Movies
# Filter only Movies and extract minutes as integers
movie_df = df[df['type'] == 'Movie'].copy()
movie_df['duration_mins'] = movie_df['duration'].str.extract(r'(\
d+)').astype(float)
avg_duration = movie_df['duration_mins'].mean()
```

```
print(f"Average Movie Duration on Netflix: {avg_duration:.2f}
minutes")
```

```
Average Movie Duration on Netflix: 99.58 minutes
```

```
# 16. Most common number of seasons for TV Shows
tv_df = df[df['type'] == 'TV Show'].copy()
tv_df['seasons'] = tv_df['duration'].str.extract(r'(\d+)').astype(int)
most_common_seasons = tv_df['seasons'].value_counts().head(1)
print("\nMost Common Number of Seasons for TV Shows:")
print(most_common_seasons)

# Plot

tv_df['seasons'].value_counts().sort_index().plot(kind='bar',
title='Distribution of Number of Seasons in TV Shows', color='orchid')
plt.xlabel('Number of Seasons')
plt.ylabel('Number of Shows')
plt.tight_layout()
plt.show()
```

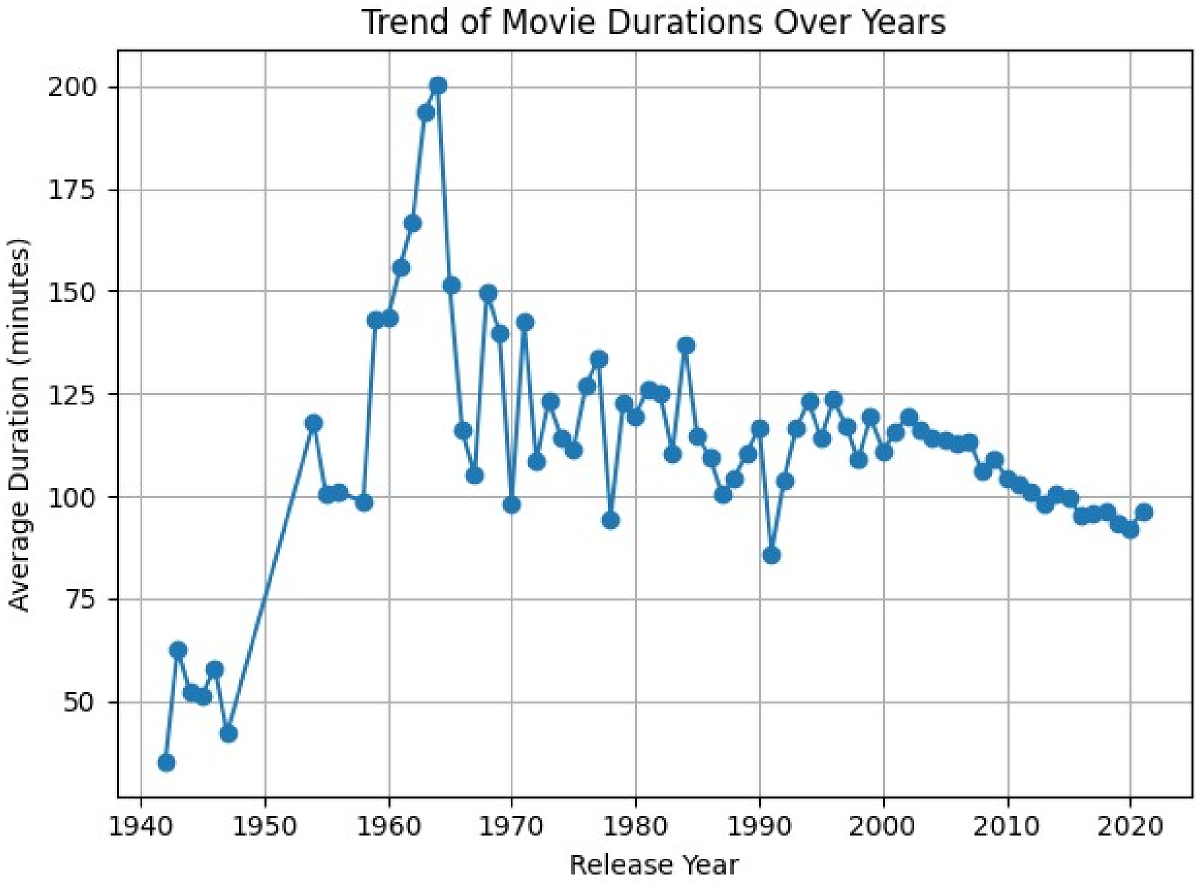
```

Most Common Number of Seasons for TV Shows:
seasons
1    1793
Name: count, dtype: int64
```



```
# 17. Trend in Movie Durations Over the Years
duration_by_year = movie_df.groupby('release_year')
['duration_mins'].mean()
print("\nAverage Movie Duration Over the Years:")
print(duration_by_year.tail(10))

# Plot
duration_by_year.plot(kind='line', marker='o', title='Trend of Movie
Durations Over Years')
plt.xlabel('Release Year')
plt.ylabel('Average Duration (minutes)')
plt.grid(True)
plt.tight_layout()
plt.show()
```

```
Average Movie Duration Over the Years:
release_year
2012    100.757225
2013     98.048889
2014    100.261364
```

```
2015    99.513854
2016    95.360182
2017    95.535248
2018    96.185137
2019    93.466035
2020    92.141199
2021    96.444043
Name: duration_mins, dtype: float64
```



content Launch Strategy

```
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
```

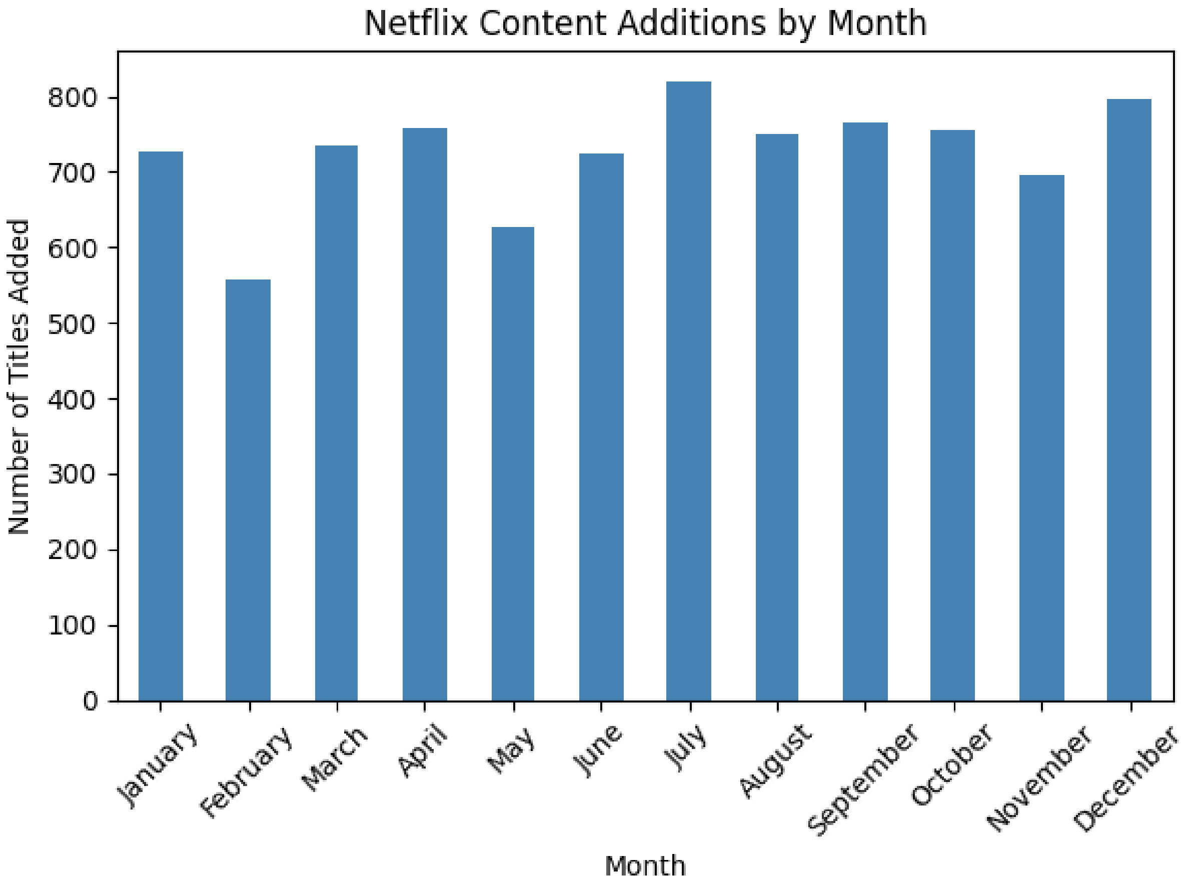
```
                                                cast        country  \
0                                                NaN  United States   
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...   South Africa   
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...            NaN   
3                                                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  
year_added  \
0  As her father nears the end of his life, filmm...  September 25, 
2021   
1  After crossing paths at a party, a Cape Town t...  September 24, 
2021   
2  To protect his family from a powerful drug lor...  September 24, 
2021   
3  Feuds, flirtations and toilet talk go down amo...  September 24, 
2021   
4  In a city of coaching centers known to train I...  September 24, 
2021   

   is_US         month_added  
0   True  September 25, 2021  
1  False  September 24, 2021  
2  False  September 24, 2021  
3  False  September 24, 2021  
4  False  September 24, 2021  
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 15 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
 12  year_added    8797 non-null   object
 13  is_US         8807 non-null   bool
 14  month_added   8797 non-null   object
dtypes: bool(1), int64(1), object(13)
memory usage: 972.0+ KB
```

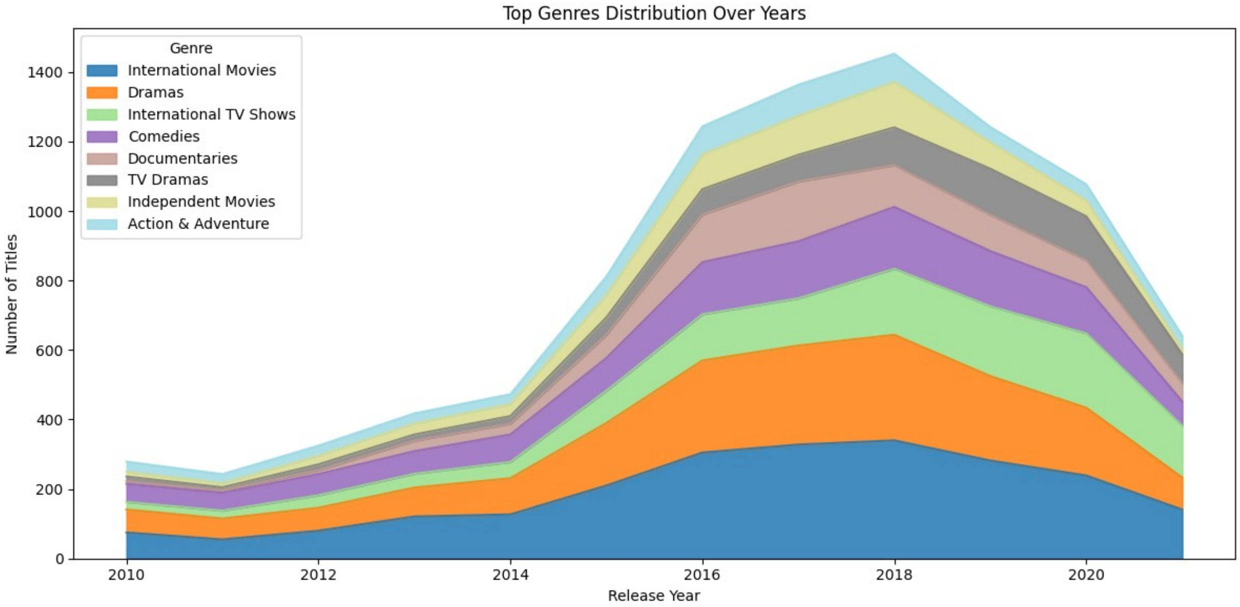
```python
# 18. In which months does Netflix add the most content?

# Convert date_added to datetime (force errors to NaT)
df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')

# Drop rows where date_added is NaT
df_valid_dates = df.dropna(subset=['date_added'])

# Extract month name
df_valid_dates['month_added'] =
df_valid_dates['date_added'].dt.month_name()

# Count additions by month
monthly_additions =
df_valid_dates['month_added'].value_counts().reindex([
    'January', 'February', 'March', 'April', 'May', 'June',
    'July', 'August', 'September', 'October', 'November', 'December'
])

print("Monthly Content Additions:")
print(monthly_additions)

# Plot
monthly_additions.plot(kind='bar', color='steelblue', title='Netflix
Content Additions by Month')
plt.xlabel('Month')
plt.ylabel('Number of Titles Added')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

```
C:\Users\LENOVO\AppData\Local\Temp\ipykernel_1228\3022191953.py:10:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  df_valid_dates['month_added'] =
df_valid_dates['date_added'].dt.month_name()

Monthly Content Additions:
month_added
January      727
February     557
March        734
April        759
May          626
June         724
July         819
August       749
September    765
October      755
November     697
December     797
Name: count, dtype: int64
```



```
# 19. Genre distribution across different years (Stacked Area Chart)

# Preprocessing
df['release_year'] = pd.to_numeric(df['release_year'],
errors='coerce')
genre_year_df = df[['release_year', 'listed_in']].dropna()
genre_year_df =
genre_year_df.assign(listed_in=genre_year_df['listed_in'].str.split(',
'))
genre_year_exploded = genre_year_df.explode('listed_in')

# Group by year and genre
genre_by_year = genre_year_exploded.groupby(['release_year',
'listed_in']).size().unstack(fill_value=0)

# Filter to recent years and top genres for clarity
recent_genre_by_year = genre_by_year.loc[genre_by_year.index >= 2010]
top_genres =
recent_genre_by_year.sum().sort_values(ascending=False).head(8).index
filtered = recent_genre_by_year[top_genres]

# Plot as stacked area chart
```

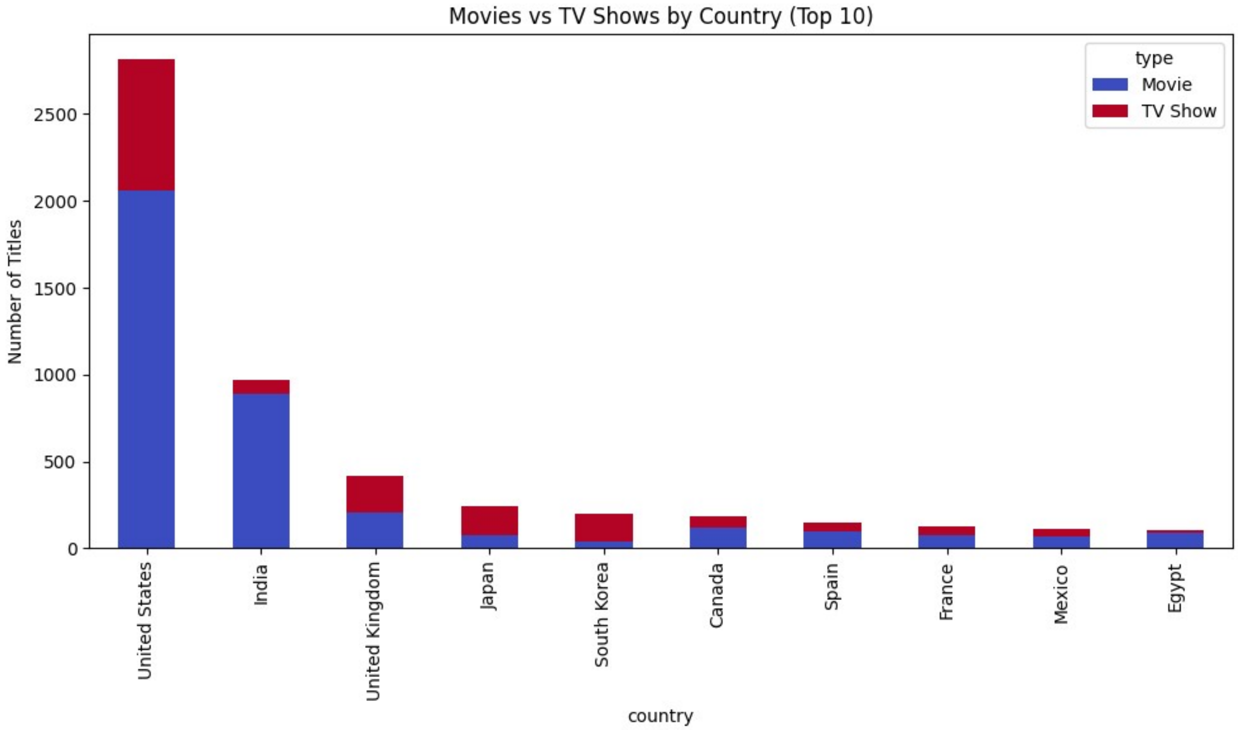
```
filtered.plot(kind='area', stacked=True, figsize=(12, 6),
cmap='tab20', alpha=0.85)
plt.title('Top Genres Distribution Over Years')
plt.xlabel('Release Year')
plt.ylabel('Number of Titles')
plt.legend(title='Genre', loc='upper left')
plt.tight_layout()
plt.show()
```



```
# 20. Countries producing the most content in each genre
genre_country_df = df[['country', 'listed_in']].dropna()
genre_country_df = genre_country_df.assign(
    country=genre_country_df['country'].str.split(', '),
    listed_in=genre_country_df['listed_in'].str.split(', ')
)

# Explode both columns
genre_country_exploded =
genre_country_df.explode('country').explode('listed_in')

# Group by genre and country
top_genre_countries = genre_country_exploded.groupby(['listed_in',
'country']).size().reset_index(name='count')

# For each genre, get the country with the highest count
top_countries_by_genre = top_genre_countries.sort_values(['listed_in',
'count'], ascending=[True, False]).groupby('listed_in').head(1)

print("\nTop Country for Each Genre:")
print(top_countries_by_genre.head(10))
```

```
Top Country for Each Genre:
                     listed_in         country  count
57          Action & Adventure   United States    404
60              Anime Features           Japan     61
66                Anime Series           Japan    143
87             British TV Shows  United Kingdom    225
135  Children & Family Movies   United States    390
141           Classic & Cult TV   United States     17
162              Classic Movies   United States     81
225                    Comedies   United States    680
277              Crime TV Shows   United States    145
294                 Cult Movies   United States     52
```

```
# Count of Movies vs TV Shows by Country
country_type = df[['country', 'type']].dropna()
country_type = country_type.groupby(['country',
'type']).size().unstack().fillna(0)

# Plot top 10 countries by total content
top_countries =
country_type.sum(axis=1).sort_values(ascending=False).head(10)
country_type.loc[top_countries.index].plot(kind='bar', stacked=True,
figsize=(10,6), colormap='coolwarm')
plt.title('Movies vs TV Shows by Country (Top 10)')
plt.ylabel('Number of Titles')
plt.tight_layout()
plt.show()
```

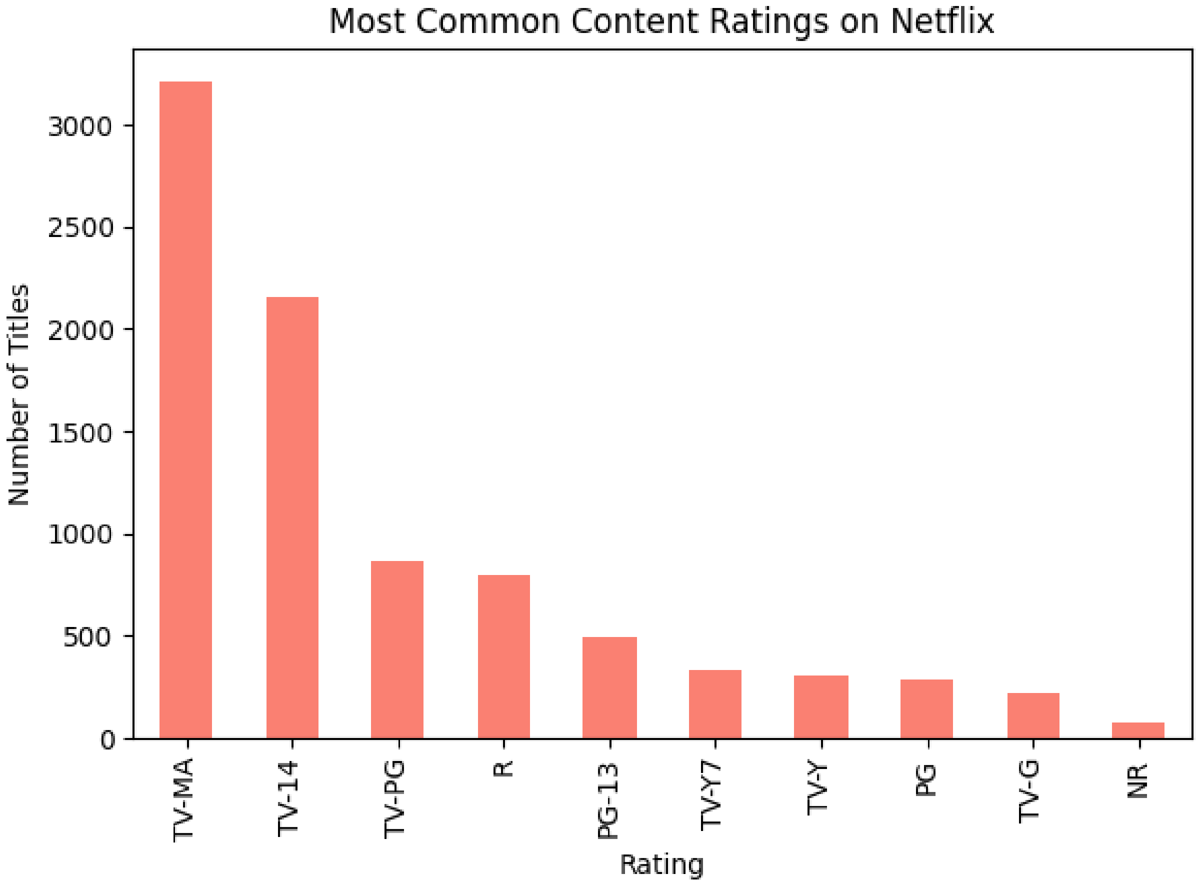


```
movie_df = df[df['type'] == 'Movie'].copy()
movie_df['duration_mins'] = movie_df['duration'].str.extract(r'(\
d+)').astype(float)
longest_movies = movie_df.sort_values('duration_mins',
ascending=False).head(10)
print("Top 10 Longest Movies on Netflix:")
print(longest_movies[['title', 'duration', 'country',
'release_year']])
```

```
Top 10 Longest Movies on Netflix:
                            title duration                country  \
4253   Black Mirror: Bandersnatch  312 min          United States
717   Headspace: Unwind Your Mind  273 min                    NaN
2491       The School of Mischief  253 min                  Egypt
2487               No Longer kids  237 min                  Egypt
2484           Lock Your Girls In  233 min                    NaN
2488              Raya and Sakina  230 min                    NaN
166   Once Upon a Time in America  229 min   Italy, United States
7932                       Sangam  228 min                  India
1019                       Lagaan  224 min  India, United Kingdom
4573                 Jodhaa Akbar  214 min                  India

      release_year
4253          2018
717           2021
2491          1973
2487          1979
```

```
2484          1982
2488          1984
166           1984
7932          1964
1019          2001
4573          2008
```

```
rating_counts = df['rating'].value_counts().head(10)
rating_counts.plot(kind='bar', color='salmon', title='Most Common
Content Ratings on Netflix')
plt.xlabel('Rating')
plt.ylabel('Number of Titles')
plt.tight_layout()
plt.show()
```

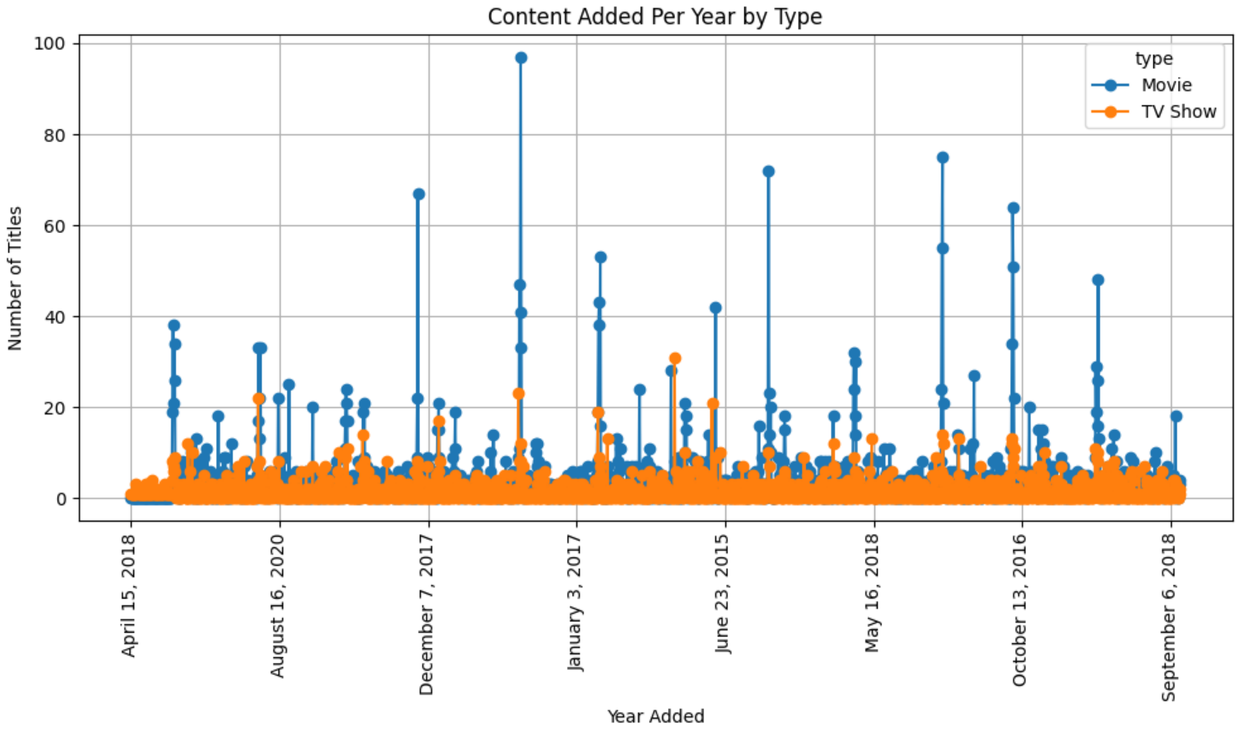


```
genre_rating_df = df[['rating', 'listed_in']].dropna()
genre_rating_df['listed_in'] =
genre_rating_df['listed_in'].str.split(', ')
genre_rating_exploded = genre_rating_df.explode('listed_in')

# Top genres per rating
top_genres_by_rating = genre_rating_exploded.groupby(['rating',
'listed_in']).size().reset_index(name='count')
top_genres_by_rating = top_genres_by_rating.sort_values(['rating',
'count'], ascending=[True, False])
print("Top Genres by Rating:")
print(top_genres_by_rating.groupby('rating').head(3))  # Top 3 per
rating
```

```
Top Genres by Rating:
       rating                 listed_in  count
0      66 min                    Movies      1
1      74 min                    Movies      1
2      84 min                    Movies      1
4           G  Children & Family Movies     33
6           G                  Comedies     11
5           G            Classic Movies      8
12      NC-17                  Comedies      2
14      NC-17        Independent Movies      2
15      NC-17      International Movies      2
27         NR      International Movies     34
24         NR                    Dramas     27
22         NR             Documentaries     22
44         PG  Children & Family Movies    195
46         PG                  Comedies    148
49         PG                    Dramas     69
66      PG-13                    Dramas    192
63      PG-13                  Comedies    168
59      PG-13        Action & Adventure    148
83          R                    Dramas    375
78          R        Action & Adventure    220
86          R        Independent Movies    193
116     TV-14      International Movies   1065
112     TV-14                    Dramas    693
117     TV-14    International TV Shows    472
153      TV-G      International Movies     54
143      TV-G  Children & Family Movies     51
148      TV-G             Documentaries     47
191     TV-MA      International Movies   1130
187     TV-MA                    Dramas    830
192     TV-MA    International TV Shows    714
232     TV-PG      International Movies    294
228     TV-PG                    Dramas    200
226     TV-PG             Documentaries    167
264      TV-Y                  Kids' TV    176
259      TV-Y  Children & Family Movies    113
271      TV-Y               TV Comedies     28
285     TV-Y7                  Kids' TV    189
277     TV-Y7  Children & Family Movies    129
295     TV-Y7               TV Comedies     54
299  TV-Y7-FV  Children & Family Movies      5
300  TV-Y7-FV                  Comedies      4
```

```
298  TV-Y7-FV        Action & Adventure      1
306        UR                    Dramas      2
307        UR      International Movies      2
308        UR           Romantic Movies      2
```

```
df['year_added'] = df['date_added']
trend_df = df.groupby(['year_added',
'type']).size().unstack().fillna(0)

# Plot trend
trend_df.plot(kind='line', marker='o', figsize=(10, 6), title='Content
Added Per Year by Type')
plt.xlabel('Year Added')
plt.ylabel('Number of Titles')
plt.grid(True)
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```


Content Added Per Year by Type
type
Movie
TV Show
Number of Titles
April 15, 2018
August 16, 2020
December 7, 2017
January 3, 2017
June 23, 2015
May 16, 2018
October 13, 2016
September 6, 2018
Year Added