IMPORT LIBRARIES

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

UPLOAD DATA

```
Netflix = pd.read_csv("/content/netflix_titles.csv")
```

SEE THE TOP 5 ROWS

```
Netflix.head(5)
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	93 min	Dramas, International Movies	After a devastating earthquake hits Mexico Cit...
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	78 min	Horror Movies, International Movies	When an army recruit is found dead, his fellow...
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer	United States	November 16, 2017	2009	PG-13	80 min	Action & Adventure, Independent Movies, Sci-	In a postapocalyptic world, rag-doll robots hi

SEE THE BOTTOM 5 ROWS

```
Netflix.tail(5)
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
7782	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car...	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	2005	TV-MA	99 min	Dramas, International Movies	When Lebanon's Civil War deprives Zozo of his ...
7783	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...
7784	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	2019	TV-MA	44 min	Documentaries, International Movies, Music & M...	In this documentary, South African rapper Nast...
7785	s7786	TV Show	Zumbo's Just Desserts	NaN	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	2019	TV-PG	1 Season	International TV Shows, Reality TV	Dessert wizard Adriano Zumbo looks for the nex...
			ZZ TOP: THAT			United						This

CHECK THE SHAPE OF DATA

```
Netflix.shape
```

```
(7787, 12)
```

ABOUT THE SIZE OF DATA

```
Netflix.size
```

```
93444
```

GET AN INFORMATION OF DATA

```
Netflix.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7787 entries, 0 to 7786
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       7787 non-null   object
 1   type          7787 non-null   object
 2   title         7787 non-null   object
 3   director      5398 non-null   object
 4   cast          7069 non-null   object
 5   country       7280 non-null   object
 6   date_added    7777 non-null   object
 7   release_year  7787 non-null   int64 
 8   rating        7780 non-null   object
 9   duration      7787 non-null   object
 10  listed_in     7787 non-null   object
 11  description   7787 non-null   object
dtypes: int64(1), object(11)
memory usage: 730.2+ KB
```

SEE THE STATISTICAL SUMMARY OF DATA

```
Netflix.describe()
```

	release_year
count	7787.000000
mean	2013.932580
std	8.757395
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2018.000000
max	2021.000000

CHECK COLUMNS

```
Netflix.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object')
```

DROP COLUMNS

```
Netflix.drop(columns=['cast', 'description'], inplace=True)
```

RENAME COLUMN

```
Netflix.rename(columns={'listed_in': 'listed_in/GENRE'}, inplace=True)
```

CHECK FOR THE NULL OR MISSING VALUES

```
Netflix.isnull().sum()
```

	0
show_id	0
type	0
title	0
director	2389
country	507
date_added	10
release_year	0
rating	7
duration	0
listed_in/GENRE	0

dtype: int64

director → 2,389 missing

country → 507 missing

date_added → 10 missing

rating → 7 missing

HANDLING NULL VALUES BY IMPUTATION

```
Netflix['rating'] = Netflix['rating'].fillna(Netflix['rating'].mode()[0])
Netflix['country'] = Netflix['country'].fillna(Netflix['country'].mode()[0])
```

```
Netflix['director'] = Netflix['director'].fillna(Netflix['country'].mode()[0])
Netflix['date_added'] = Netflix['date_added'].fillna(Netflix['date_added'].mode()[0])
```

CHECK AGAIN, NOW IT CONTAINS NO NULLS

```
Netflix.isnull().sum()
```

	0
show_id	0
type	0
title	0
director	0
country	0
date_added	0
release_year	0
rating	0
duration	0
listed_in/GENRE	0

dtype: int64

NON-GRAPHICAL ANALYSIS

```
Netflix['show_id'].value_counts()
```

show_id	count
s7787	1
s1	1
s2	1
s3	1
s4	1
...	...
s16	1
s15	1
s14	1
s13	1
s12	1

7787 rows × 1 columns

dtype: int64

How many Movies vs TV Shows are in the data?

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

type	count
Movie	5377
TV Show	2410

dtype: int64

There are **5377 Movies** and **2410 TV shows** in the Netflix.

```
Netflix['title'].value_counts()
```

title	count
ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	1
3%	1
07:19	1
23:59	1
9	1
...	...
Oct-01	1
3022	1
2,215	1
1994	1
1983	1

7787 rows × 1 columns

dtype: int64

```
Netflix['director'].value_counts()
```

director	count
United States	2389
Raúl Campos, Jan Suter	18
Marcus Raboy	16
Jay Karas	14
Cathy Garcia-Molina	13
...	...
Jonathan Helpert	1
Greg Kohs	1
Jacob Schwab	1
Serge Ou	1
Michael Gallagher	1

4050 rows × 1 columns

dtype: int64

Which movies/TV shows have no director listed?

```
no_director = Netflix[Netflix['director'].isna()]
no_director
```

show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in/GENRE

There is No such movie/tv shows have no director listed.

Which countries appear most frequently in the data?

```
Netflix['country'].value_counts()
```

country	count
United States	3062
India	923
United Kingdom	397
Japan	226
South Korea	183
...	...
Germany, United States, United Kingdom, Canada	1
Peru, United States, United Kingdom	1
Saudi Arabia, United Arab Emirates	1
United Kingdom, France, United States, Belgium	1
France, Norway, Lebanon, Belgium	1

681 rows × 1 columns

dtype: int64

United States is the country which is appear most frequently in the Netflix data.

```
Netflix['release_year'].value_counts()
```

release_year	count
2018	1121
2017	1012
2019	996
2016	882
2020	868
...	...
1966	1
1925	1
1964	1
1947	1
1959	1

73 rows × 1 columns

dtype: int64

What is the earliest and latest release year of the shows/movies?

```
earliest_year = Netflix['release_year'].min()
latest_year = Netflix['release_year'].max()
```

```
earliest_year
```

```
1925
```

```
latest_year
```

```
2021
```

earliest release year of the shows/movies 1925 & latest release year of the shows/movies 2021

```
Netflix['rating'].value_counts()
```

rating	count
TV-MA	2870
TV-14	1931
TV-PG	806
R	665
PG-13	386
TV-Y	280
TV-Y7	271
PG	247
TV-G	194
NR	84
G	39
TV-Y7-FV	6
UR	5
NC-17	3

dtype: int64

Highest rating for TV-MA which is 2970 and the lowest rating which is for NC-17 is 3 only.

What are the different rating categories and how many titles fall under each?

```
type_rating_counts = Netflix.groupby(['title', 'rating']).size().reset_index(name='count')
type_rating_counts.head(30)
```

	title	rating	count
0	#Alive	TV-MA	1
1	#AnneFrank - Parallel Stories	TV-14	1
2	#FriendButMarried	TV-G	1
3	#FriendButMarried 2	TV-G	1
4	#Roxy	TV-14	1
5	#Rucker50	TV-PG	1
6	#Selfie	TV-MA	1
7	#Selfie 69	TV-MA	1
8	#blackAF	TV-MA	1
9	#cats_the_mewvie	TV-14	1
10	#realityhigh	TV-14	1
11	'89	TV-PG	1
12	(T)ERROR	NR	1
13	(Un)Well	TV-MA	1
14	07:19	TV-MA	1
15	1 Chance 2 Dance	TV-PG	1
16	1 Mile to You	TV-14	1
17	10 Days in Sun City	TV-14	1
18	10 jours en or	TV-14	1
19	10,000 B.C.	PG-13	1
20	100 Days My Prince	TV-14	1
21	100 Days Of Solitude	TV-MA	1
22	100 Humans	TV-14	1
23	100 Meters	TV-MA	1
24	100 Things to do Before High School	TV-Y	1
25	100 Years: One Woman's Fight for Justice	TV-14	1
26	100% Halal	TV-14	1
27	100% Hotter	TV-14	1
28	1000 Rupee Note	TV-14	1
29	12 ROUND GUN	TV-MA	1

It contains titles across multiple rating categories. **The largest category is TV-MA, which has the highest number of titles, followed by TV-14.** On the other hand, the **smallest categories are TV-G and TV-Y7, which have the least number of titles.** This shows that the catalog is dominated by mature-rated content, while family/kids-oriented content forms only a small portion."

```
Netflix['duration'].value_counts()
```

duration	count
1 Season	1608
2 Seasons	382
3 Seasons	184
90 min	136
93 min	131
...	...
36 min	1
201 min	1
253 min	1
203 min	1
191 min	1

216 rows × 1 columns

dtype: int64

```
Netflix['listed_in/GENRE'].value_counts()
```

listed_in/GENRE	count
Documentaries	334
Stand-Up Comedy	321
Dramas, International Movies	320
Comedies, Dramas, International Movies	243
Dramas, Independent Movies, International Movies	215
...	...
Crime TV Shows, International TV Shows, TV Sci-Fi & Fantasy	1
Docuseries, Science & Nature TV, TV Action & Adventure	1
British TV Shows, Classic & Cult TV, Kids' TV	1
Docuseries, TV Sci-Fi & Fantasy	1
Children & Family Movies, Dramas, Music & Musicals	1

492 rows × 1 columns

dtype: int64

It shows 492 unique genre categories. The largest category is **Documentaries, with 334 titles**, followed closely by **Stand-Up Comedy (321 titles)**and Dramas, International Movies (320 titles)**bold text** Other popular categories include Comedies, Dramas, International Movies (243 titles) and Dramas, Independent Movies, International Movies (215 titles). On the other hand, a large number of niche categories (such as **Crime TV Shows, International TV Shows, TV Sci-Fi & Fantasy or Docuseries, TV Sci-Fi & Fantasy) appear only once** This indicates that Most of the catalog is concentrated in a few big genres, but the platform also offers lots of rare, niche genres that give viewers more diverse options.

How many shows/movies were added per year on Netflix (based on date_added)?

```
# Convert all to string, strip spaces, then convert to datetime
Netflix['date_added'] = pd.to_datetime(Netflix['date_added'].astype(str).str.strip(), errors='coerce')
```

```
# Extract the year
Netflix['year_added'] = Netflix['date_added'].dt.year
```

```
# Number of titles added per year
added_per_year = Netflix['year_added'].value_counts().sort_index()
```

```
added_per_year
```

year_added	count
2008	2
2009	2
2010	1
2011	13
2012	3
2013	11
2014	25
2015	88
2016	443
2017	1225
2018	1685
2019	2153
2020	2019
2021	117

dtype: int64

"Netflix added only a handful of **titles before 2015**, but the number of additions grew rapidly after that. The **highest growth occurred in 2019 (2,153 titles)**, followed by 2020 (2,019 titles) and 2018 (1,685 titles). The s**mallest years were 2008 and 2009 with just 2 titles each.** This shows a period of **rapid expansion** between** 2016–2020,** followed by a **decline in 2021 (117 titles)**."

Which country contributes more TV Shows vs Movies?

```
Netflix['country'] = Netflix['country'].str.split(',') # First, split the 'country' column into lists (converts into list)
```

```
Netflix_exploded = Netflix.explode('country')  # Explode the DataFrame so each country has its own row
```

```
Netflix_exploded['country'] = Netflix_exploded['country'].str.strip() # Remove extra spaces
```

```
# Now, group by on country and type to geeting the contribution to the tv shows or movies with respect of countries..
country_type_counts = Netflix_exploded.groupby(['country', 'type']).size().reset_index(name='count')
country_type_counts
```

	country	type	count
0		Movie	4
1	Afghanistan	Movie	1
2	Albania	Movie	1
3	Algeria	Movie	2
4	Angola	Movie	1
...	...	...	...
171	Venezuela	Movie	3
172	Vietnam	Movie	5
173	West Germany	Movie	3
174	West Germany	TV Show	2
175	Zimbabwe	Movie	3

176 rows × 3 columns

Across 176 countries, Movies make up the majority of content contributions, while TV Shows are much less common and only appear in a smaller set of countries.

The only country contributing TV Shows is:

West Germany ⟶ 2 TV Shows

All other listed countries in your sample only contributed Movies.

GRAPHICAL ANALYSIS

Are there patterns between release year and rating (e.g., older films rated differently than recent ones)?

```
netflix_counts=Netflix.groupby(['release_year', 'rating']).size().reset_index(name='count')
```

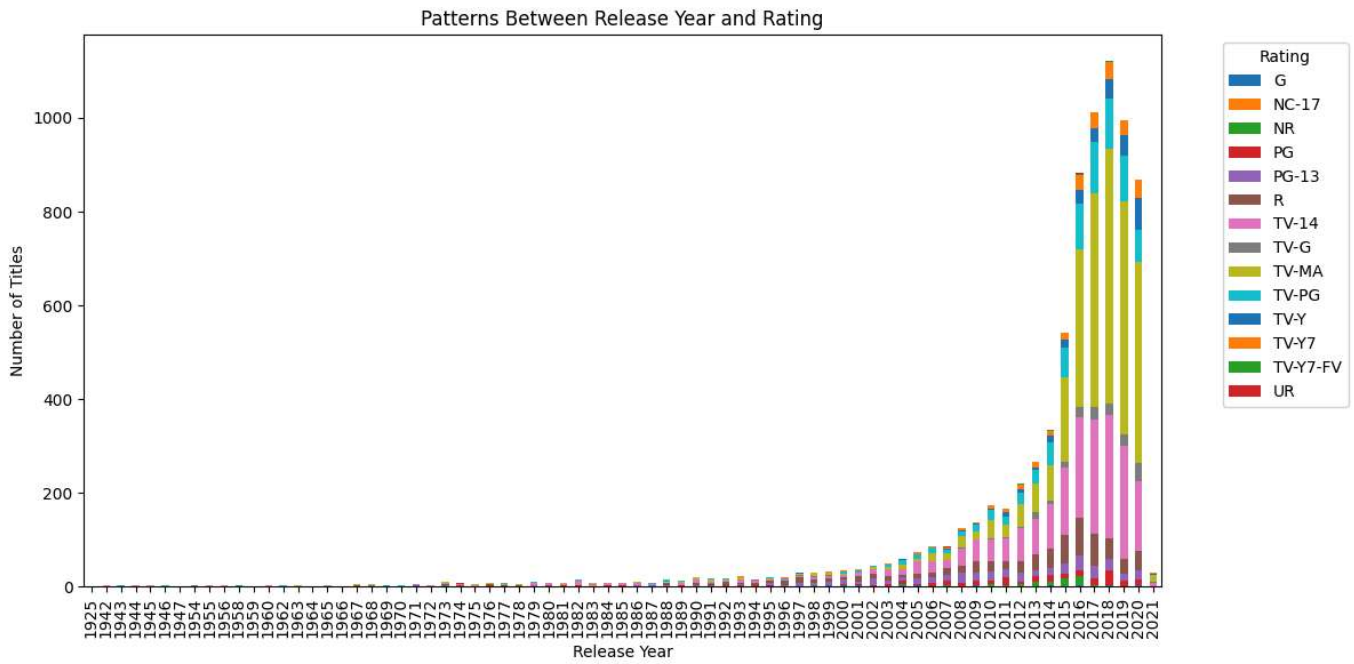
```
pivot_table = Netflix.pivot_table(index='release_year', columns='rating', aggfunc='size', fill_value=0) # aggfunc for counts,mean,m
pivot_table
```

rating	G	NC-17	NR	PG	PG-13	R	TV-14	TV-G	TV-MA	TV-PG	TV-Y	TV-Y7	TV-Y7-FV	UR
release_year														
1925	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1942	0	0	0	0	0	0	2	0	0	0	0	0	0	0
1943	0	0	0	0	0	0	0	0	0	3	0	0	0	0
1944	0	0	0	0	0	0	2	0	0	1	0	0	0	0
1945	0	0	0	0	0	0	2	0	1	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2017	1	0	2	15	26	68	245	26	455	111	29	34	0	0
2018	2	1	1	30	26	43	263	25	544	106	42	37	1	0
2019	0	0	0	12	15	33	240	24	499	97	44	32	0	0
2020	2	0	0	14	19	41	150	39	427	70	67	39	0	0
2021	0	0	0	0	0	1	7	2	14	2	2	3	0	0

73 rows × 14 columns

```
pivot_df = netflix_counts.pivot_table(
    index="release_year",
    columns="rating",
    values="count",
    aggfunc="sum",
    fill_value=0
)
```

```
pivot_df.plot(kind="bar", figsize=(12,6))

plt.title("Patterns Between Release Year and Rating")
plt.xlabel("Release Year")
plt.ylabel("Number of Titles")
plt.legend(title="Rating", bbox_to_anchor=(1.05, 1), loc='upper left')  # legend outside
plt.tight_layout()
plt.show()
```



There are patterns between release year and rating.**Older films**(1920s–1950s) are mostly **rated TV-PG and TV-14**, while more recent releases **(2000s–2020s) show greater diversity in ratings**, with a sharp rise in TV-MA (mature content) alongside **smaller shares of family-oriented ratings like TV-Y, TV-Y7, and TV-G.** This indicates a shift from mainly moderate ratings in older films to a wider spread with more mature content in modern releases."

There is a clear trend that older films are mostly TV-PG/TV-14, while modern content is more diverse, with a notable increase in TV-MA titles.

Which countries mostly release R-rated movies?

```
r_movies = Netflix[(Netflix['type'] == 'Movie') & (Netflix['rating'] == 'R')].copy()
r_movies
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in/GENRE	year_added
2	s3	Movie	23:59	Gilbert Chan	[Singapore]	2018-12-20	2011	R	78 min	Horror Movies, International Movies	2018
7	s8	Movie	187	Kevin Reynolds	[United States]	2019-11-01	1997	R	119 min	Dramas	2019
14	s15	Movie	3022	John Suits	[United States]	2020-03-19	2019	R	91 min	Independent Movies, Sci-Fi & Fantasy, Thrillers	2020
17	s18	Movie	22-Jul	Paul Greengrass	[Norway, Iceland, United States]	2018-10-10	2018	R	144 min	Dramas, Thrillers	2018
65	s66	Movie	13 Sins	Daniel Stamm	[United States]	2019-01-13	2014	R	93 min	Horror Movies, Thrillers	2019
...	...	...	...	...	...	...	...	...	...	...	...
7710	s7711	Movie	Yes, God, Yes	Karen Maine	[United States]	2020-10-22	2020	R	78 min	Comedies, Dramas, Independent Movies	2020
7736	s7737	Movie	Young Adult	Jason Reitman	[United States]	2019-11-20	2011	R	94 min	Comedies, Dramas, Independent Movies	2019
7758	s7759	Movie	Zack and Miri Make a Porno	Kevin Smith	[United States]	2018-10-01	2008	R	101 min	Comedies, Independent Movies, Romantic Movies	2018
7774	s7775	Movie	Zodiac	David Fincher	[United States]	2019-11-20	2007	R	158 min	Cult Movies, Dramas, Thrillers	2019

```
r_count = r_movies['country'].value_counts()
r_count
```

country	count
[United States]	365
[United Kingdom]	31
[United Kingdom, United States]	20
[Canada]	16
[United States, United Kingdom]	11
...	...
[Canada, United States, India, United Kingdom]	1
[Switzerland, United Kingdom, United States]	1
[United States, United Kingdom, Germany]	1
[United Kingdom, Belgium]	1
[United States, Canada, United Kingdom]	1

156 rows × 1 columns

dtype: int64

The dataset shows which countries are associated with producing R-rated movies.

The United States dominates with 365 titles, far more than any other country.

Other significant contributors include t**he United Kingdom (31 titles)**,**United Kingdom–United States collaborations (20), and Canada (16)**.

Smaller contributions come from multi-country collaborations (e.g., United States + United Kingdom, Canada + United States + India + United Kingdom, etc.), each with only 1 title.

- R-rated movies are overwhelmingly from the United States, with the UK and Canada playing smaller but notable roles, while multi-country productions make up only a handful.

bi or multi v graph

"Which genres dominate the Netflix catalog, and how do niche categories (like Independent Movies, Children & Family Movies, and Romantic Movies) compare to major ones like International Movies and Dramas?"

```
genres = Netflix['listed_in/GENRE'].dropna().str.split(',').explode().str.strip()
genres
```

	listed_in/GENRE
0	International TV Shows
0	TV Dramas
0	TV Sci-Fi & Fantasy
1	Dramas
1	International Movies
...	...
7784	Music & Musicals
7785	International TV Shows
7785	Reality TV
7786	Documentaries
7786	Music & Musicals

17071 rows × 1 columns

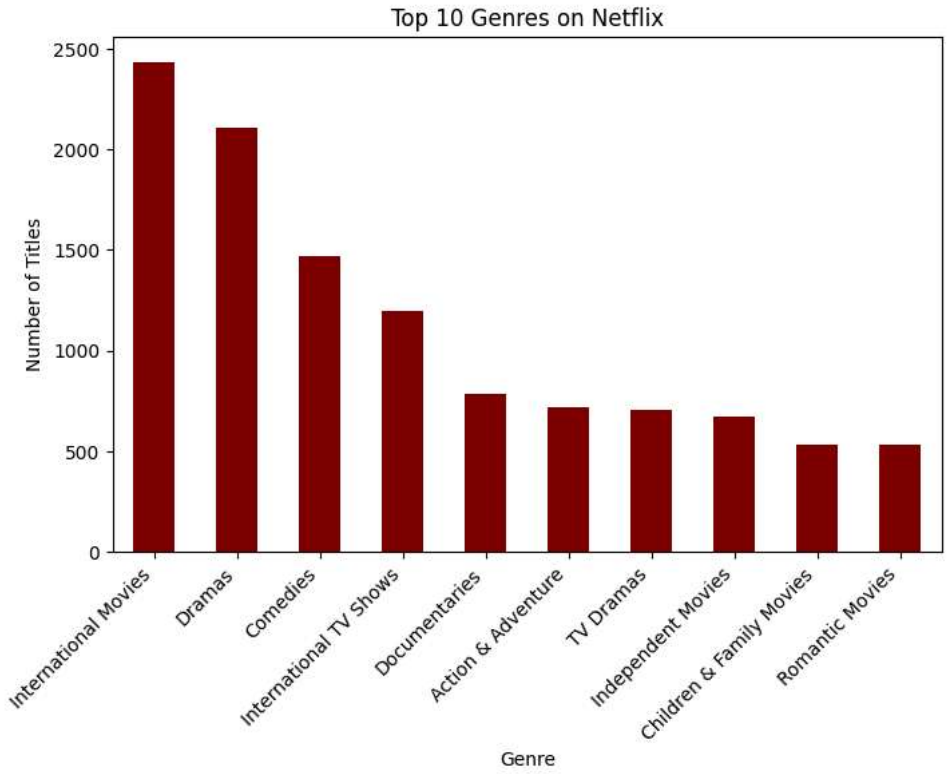
dtype: object

```
top_genres = genres.value_counts().head(10)
top_genres
```

listed_in/GENRE	count
International Movies	2437
Dramas	2106
Comedies	1471
International TV Shows	1199
Documentaries	786
Action & Adventure	721
TV Dramas	704
Independent Movies	673
Children & Family Movies	532
Romantic Movies	531

dtype: int64

```
plt.figure(figsize=(8,5))
top_genres.plot(kind='bar', color='maroon')
plt.title('Top 10 Genres on Netflix')
plt.xlabel('Genre')
plt.ylabel('Number of Titles')
plt.xticks(rotation=45, ha='right')
plt.show()
```



Insights:

The datas highlights which genres have the most titles.

International Movies dominate with 2,437 titles, followed by Dramas (2,106) and Comedies (1,471).

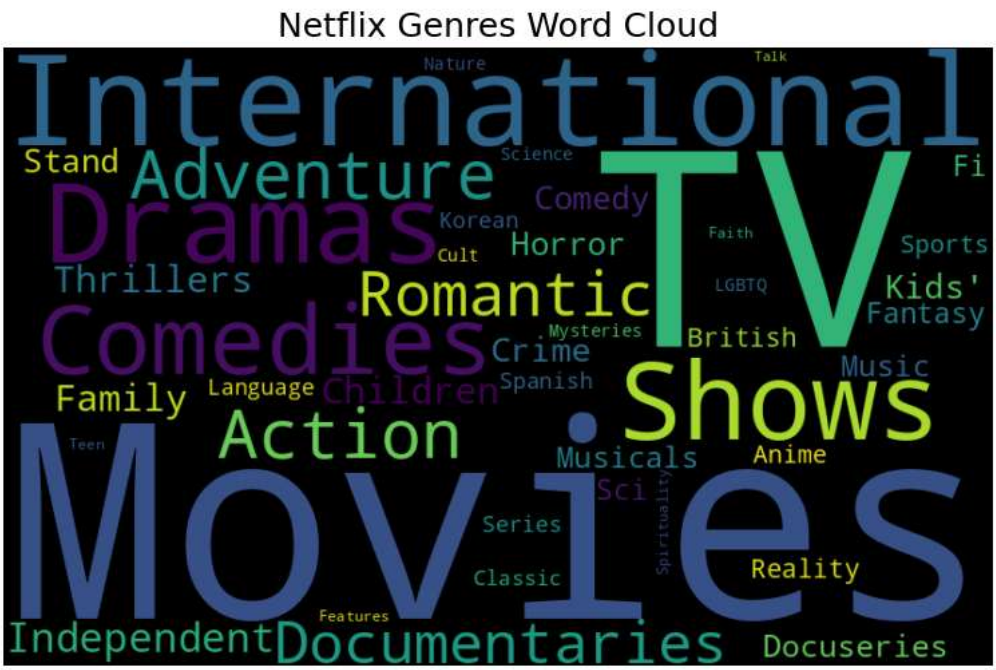
Other strong categories include International TV Shows (1,199), Documentaries (786), Action & Adventure (721), and TV Dramas (704).

Mid-level genres include Independent Movies (673), Children & Family Movies (532), and Romantic Movies (531).

This is heavily dominated by International Movies, Dramas, and Comedies, while family, romance, and niche categories make up smaller shares.

Netflix Genres Word Cloud

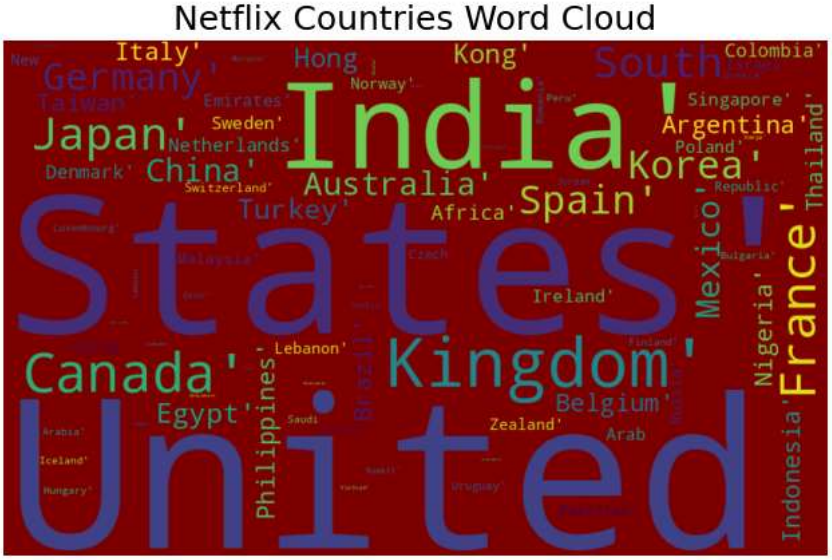
```
from wordcloud import WordCloud
text = " ".join(genres.tolist()) #Pandas Series into a Python list.
wc = WordCloud(
    width=800,
    height=500,
    background_color="black",
    collocations=False,   # don't join words like "Science" and "Fiction" unless they appear together,False → every word is treated
    max_words=150
).generate(text)


plt.figure(figsize=(12,6))
plt.imshow(wc, interpolation='bilinear')
plt.axis('off')
plt.title('Netflix Genres Word Cloud', fontsize=18)
plt.show()
```



Netflix Countries Word Cloud

"Which countries appear most frequently in Netflix titles?"

```
from wordcloud import WordCloud
text = " ".join(countries.tolist()) #Pandas Series into a Python list.
wc = WordCloud(
    width=800,
    height=500,
    background_color="maroon",
    collocations=False,   # don't join words like "Science" and "Fiction" unless they appear together,False → every word is treated
    max_words=150
).generate(text)


plt.figure(figsize=(10,5))
plt.imshow(wc, interpolation='bilinear')
plt.axis('off')
plt.title('Netflix Countries Word Cloud', fontsize=18)
plt.show()
```



Netflix Directors Word Cloud

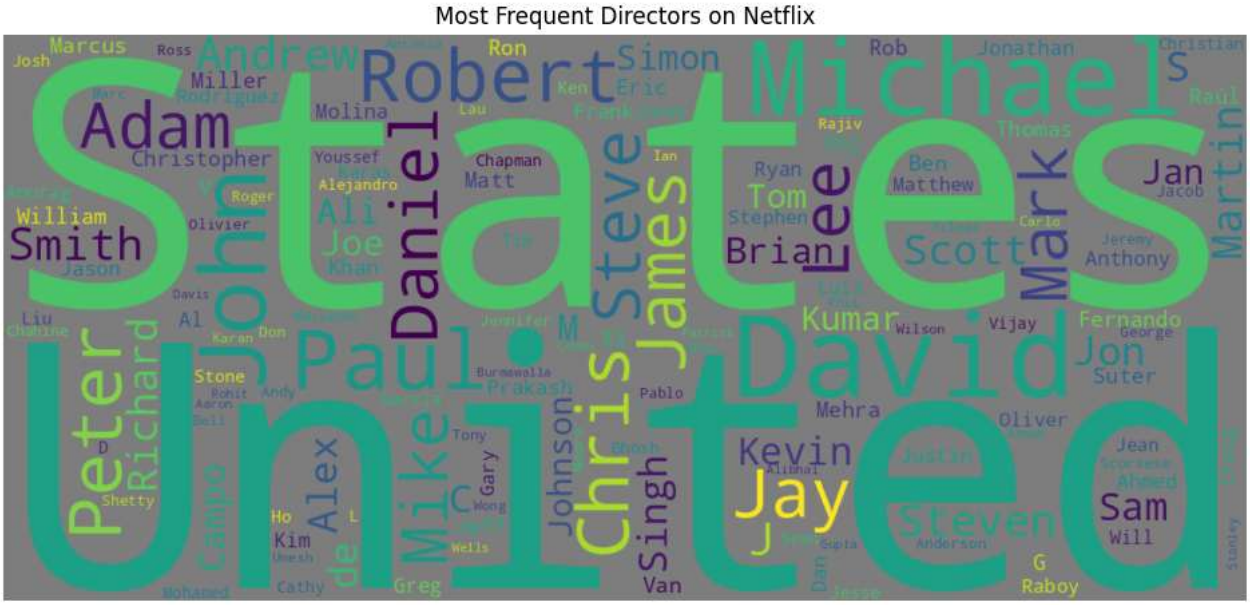
Which directors appear most frequently in Netflix content?

```
director_text = " ".join(Netflix['director'].dropna().tolist())
```

```
'United States Jorge Michel Grau Gilbert Chan Shane Acker Robert Luketic Serdar Akar Yasir Al Yasiri Kevin Reynolds Shravan Kumar Vi
kram Bhatt Zak Hilditch United States Diego Enrique Osorno Nottapon Boonprakob John Suits Kunle Afolayan United States Paul Greengra
ss Swapnaneel Jayakar United States Onir Vijay Milton Santwana Bardoloi Atanu Ghosh United States Lyric R. Cabral, David Felix Sutcl
iffe United States Cho Il Sabina Fedeli, Anna Migotto United States Michael Margolis Rako Prijanto Rako Prijanto Fernando Lebrija Mi
chael Kennedy Robert McCullough Jr. Cristina Jacob Cristina Jacob United States Frank Ariza Muharrem Gülmez Óskar Thór Axelsson Ozan
Açıktan Kenneth Gyang Karyn Kusama United States Adam Deyoe Leif Tilden Adze Ugah Nicolas Brossette Roland Emmerich United States Un
ited States United States Marcel Barrena United States Melinda Janko Jastis Arimba United States Shrihari Sathe Sam Upton United Sta
```

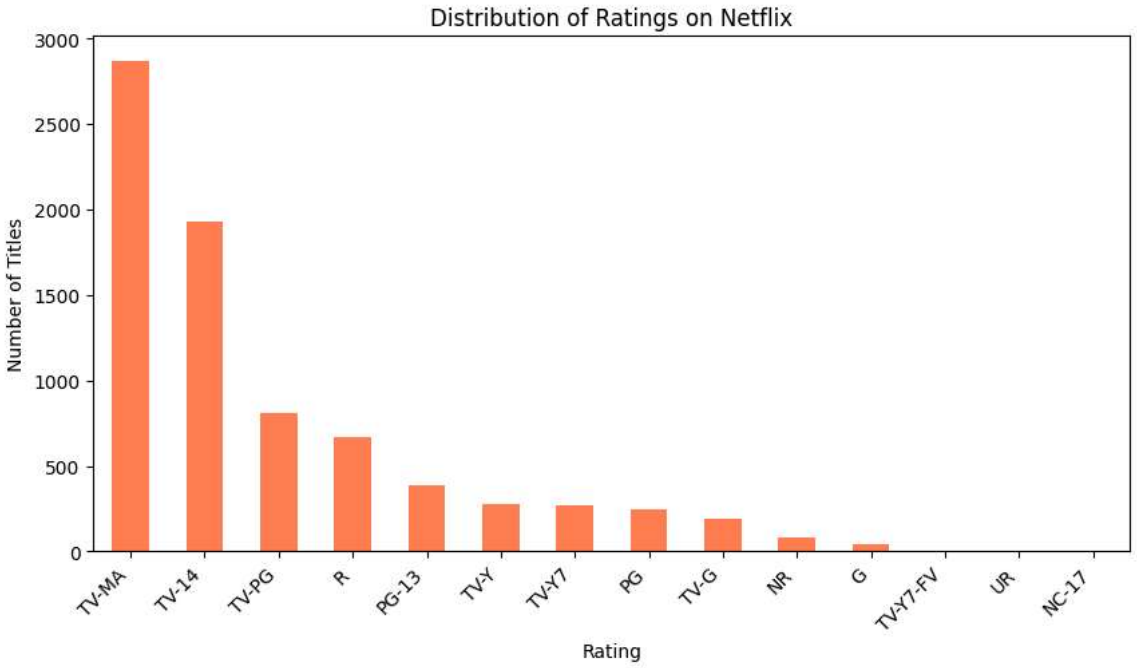
```
wc_director = WordCloud(
    width=1100, height=500,
    background_color="grey",
     collocations=False,
    max_words=150
).generate(director_text)
```

```
plt.figure(figsize=(12,6))
plt.imshow(wc_director, interpolation='bilinear')
plt.axis('off')
plt.title('Most Frequent Directors on Netflix', fontsize=12)
plt.show()
```



What is the distribution of content ratings (e.g., TV-MA, TV-14, PG-13, R) on Netflix?

```
rating_counts = Netflix['rating'].value_counts()

# Plot bar chart
plt.figure(figsize=(10,5))
rating_counts.plot(kind='bar', color='coral')

plt.title('Distribution of Ratings on Netflix')
plt.xlabel('Rating')
plt.ylabel('Number of Titles')
plt.xticks(rotation=45, ha='right')
plt.show()
```



Insights:

TV-MA dominates → The largest share of Netflix content is rated TV-MA, showing Netflix leans heavily toward mature/adult audiences.

TV-14 is the second highest → A large number of shows/movies are teen-friendly, suggesting Netflix also targets younger audiences.

Theatrical ratings like R, PG-13, PG are smaller → Netflix has fewer traditional movie ratings compared to TV-style ratings.

Very few G/TV-G titles → Netflix has limited children/family-safe content compared to mature content.

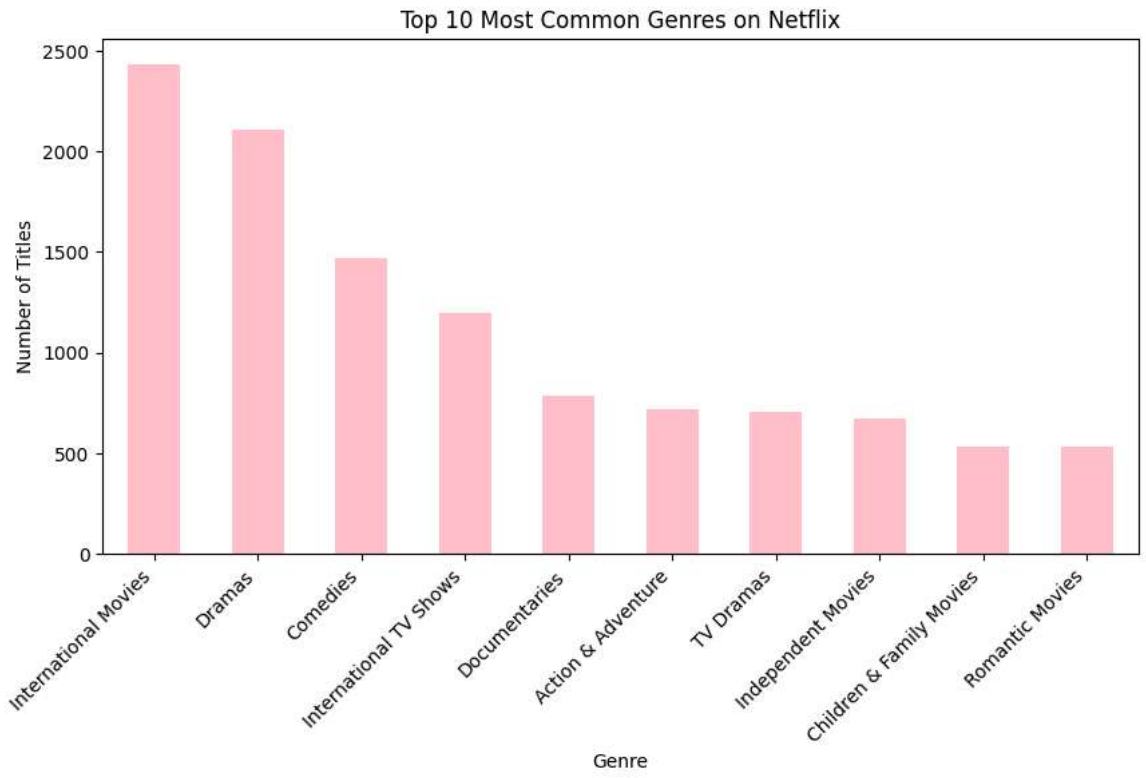
Imbalance in categories → The chart highlights that Netflix invests more in mature and teen content than in all-age categories.

Netflix content is dominated by mature (TV-MA) and teen (TV-14) ratings, while family-friendly categories make up a much smaller share.”

Which genres dominate Netflix content, and how do they compare in frequency?

```
#genres = Netflix['listed_in/GENRE'].dropna().str.split(',').explode().str.strip()
#top_genres = genres.value_counts().head(10)
```

```
plt.figure(figsize=(10,5))
top_genres.plot(kind='bar', color='pink')

plt.title('Top 10 Most Common Genres on Netflix')
plt.xlabel('Genre')
plt.ylabel('Number of Titles')
plt.xticks(rotation=45, ha='right')
plt.show()
```



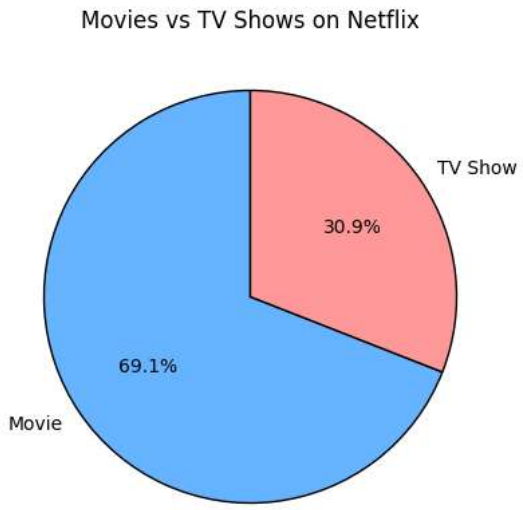
International Movies are dominant → Suggests Netflix strongly focuses on global, cross-border content.

Dramas and **Comedies are very popular** → These broad genres appeal to a wide audience and form Netflix's backbone.

International TV Shows and Documentaries are lesser as compare to them → Show Netflix's investment in them**

What proportion of Netflix content is Movies compared to TV Shows?

```
type_counts = Netflix['type'].value_counts()

# Plot pie chart
plt.figure(figsize=(5,5))
type_counts.plot(
    kind='pie',
    autopct='%1.1f%%',       # show percentage values
    startangle=90,           # rotate pie so it starts at the top
    colors=['#66b3ff','#ff9999'],  # custom colors
    wedgeprops={'edgecolor':'black'} # outline for clarity
)

plt.title('Movies vs TV Shows on Netflix')
plt.ylabel('')  # remove y-label for cleaner look
plt.show()
```



Insights:

**Movies dominate **the catalog — the pie chart will show a much larger slice for Movies compared to TV Shows.

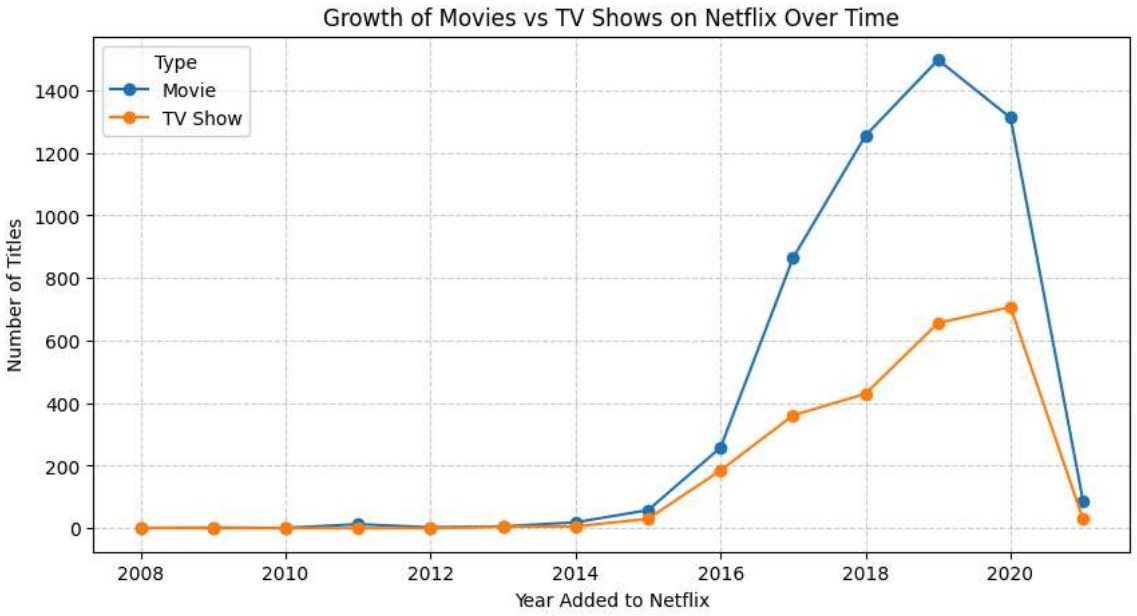
Typically, Netflix has around **70% Movies** vs **30% TV Shows** (your chart should reflect something close to this).

TV Shows form a smaller share — although fewer in number, Netflix invests heavily in TV shows because they drive longer viewer engagement (binge-watching, multiple seasons).

This pie chart shows that Movies make up the majority of Netflix's catalog, while TV Shows form a smaller but strategically important portion focused on retaining subscribers through serialized storytelling.

How has Netflix content (Movies vs TV Shows) grown over time?

```
#Netflix['date_added'] = pd.to_datetime(Netflix['date_added'], errors='coerce')
#Netflix['year_added'] = Netflix['date_added'].dt.year

# Group by year and type
year_type = Netflix.groupby(['year_added', 'type']).size().unstack(fill_value=0)

# Plot line chart
year_type.plot(kind='line', figsize=(10,5), marker='o')
plt.title('Growth of Movies vs TV Shows on Netflix Over Time')
plt.xlabel('Year Added to Netflix')
plt.ylabel('Number of Titles')
plt.grid(True, linestyle='--', alpha=0.6)
plt.legend(title='Type')
plt.show()
```



Insights:

Slow start

Before 2015, **very few Movies or TV Shows** were added each year. Netflix's catalog was still small.

Rapid growth phase

From 2015 onward, there is a **sharp rise in both Movies and TV Shows**.

The biggest jump happens between 2016–2020, when Netflix massively expanded its library.

Movies vs TV Shows trend

Movies consistently outnumber TV Shows each year.

However, **TV Shows grew faster after 2016**, showing Netflix's push into series production.

Peak years

Around **2018–2020**,** both Movies and TV Shows hit their highest numbers** — reflecting Netflix's global expansion and aggressive content strategy.

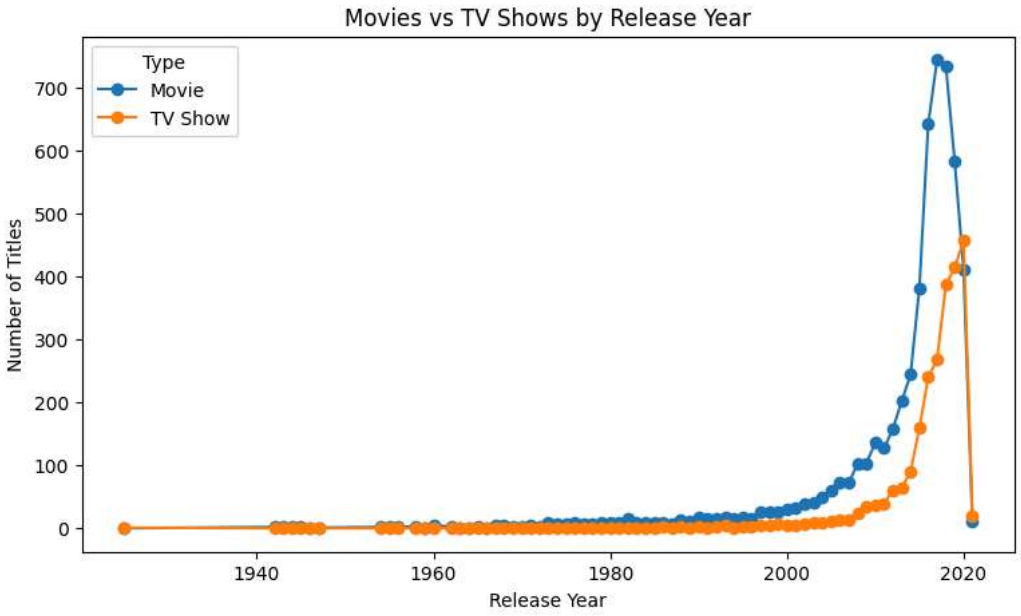
Recent decline

After 2020, the number of new additions drops (fewer Movies and TV Shows added in 2021). This could be due to COVID-19 production delays or Netflix shifting toward quality over quantity.

Netflix's catalog grew slowly at first, then exploded between 2016–2020, with Movies always leading but TV Shows gaining momentum. After 2020, the pace of new content slowed down.

How does the number of movies vs TV shows vary by release year?

```
year_type_counts = Netflix.groupby(['release_year', 'type']).size().unstack(fill_value=0)

# Plot
year_type_counts.plot(kind='line', figsize=(9,5), marker='o', color=['#1f77b4', '#ff7f0e'] )
plt.title('Movies vs TV Shows by Release Year')
plt.xlabel('Release Year')
plt.ylabel('Number of Titles')
plt.legend(title='Type')
plt.show()
```



Early decades (before 1980s)

Almost no TV Shows are present, and only a small number of Movies appear.

Netflix's older catalog is **very limited**.

1980s–2000s

Gradual growth in Movies, **but still very few TV Shows**.

This suggests Netflix has fewer older TV series in its library compared to films.

Post-2000s

Significant** rise in both Movies and TV Shows.**

The growth of TV Shows becomes noticeable mainly after 2010, reflecting the global boom in serialized streaming content.

Peak release years

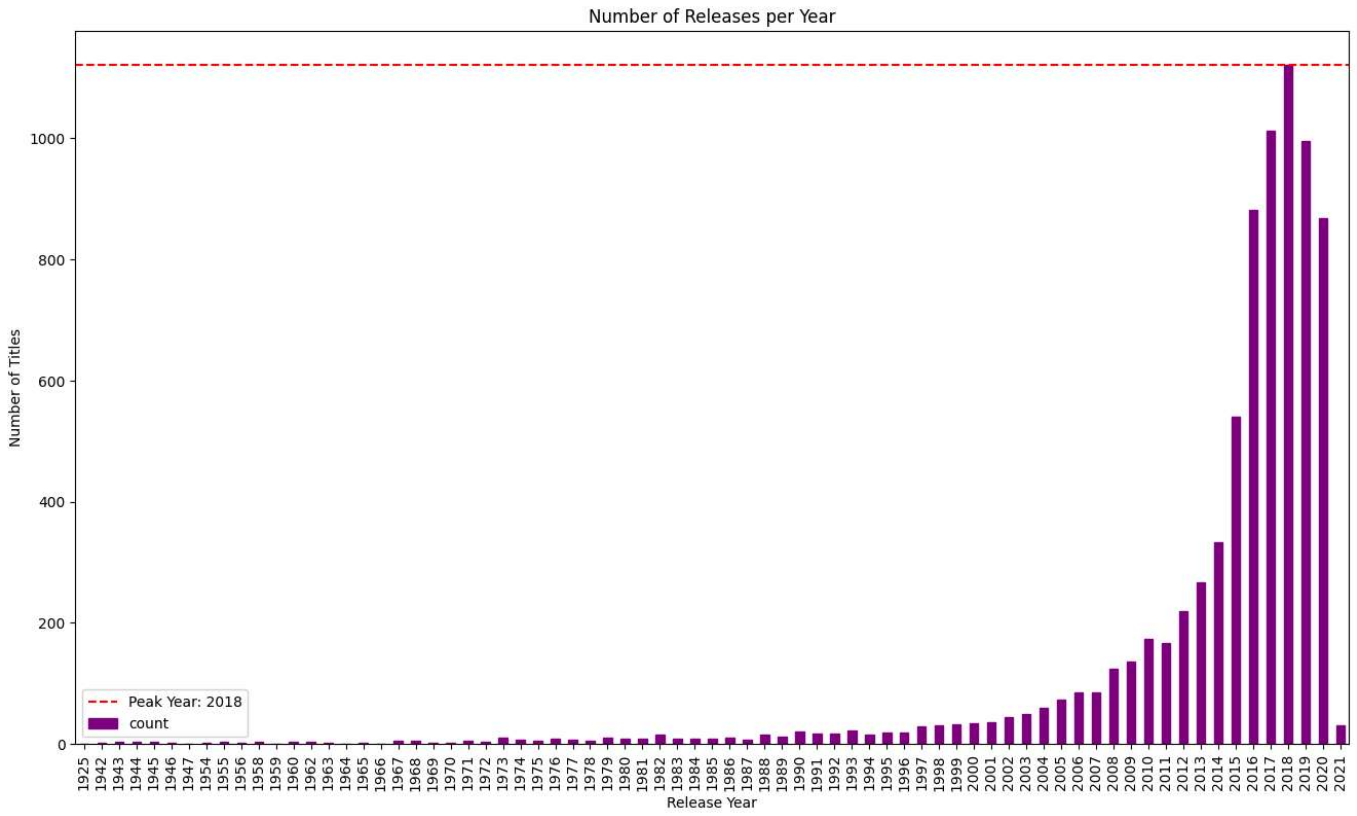
2**015–2020 release years dominate both Movies and TV Shows**, showing Netflix's focus on modern, recent content.

Movies remain more numerous overall, but the gap between Movies and TV Shows narrows in the 2010s, indicating Netflix's strategy to expand its TV Show catalog.

Most Netflix content comes from recent release years (2010 onward), with Movies consistently leading but TV Shows catching up strongly in the last decade.

Which year had the highest number of releases among the given titles in code?

```
year_counts = Netflix['release_year'].value_counts().sort_index()
max_year = year_counts.idxmax()
max_count = year_counts.max()

#plot
import matplotlib.pyplot as plt

plt.figure(figsize=(16,9))
year_counts.plot(kind='bar', color='purple', edgecolor='purple')
plt.axhline(max_count, color='red', linestyle='--', label=f'Peak Year: {max_year}')
plt.title('Number of Releases per Year')
plt.xlabel('Release Year')
plt.ylabel('Number of Titles')
plt.legend()
plt.show()
```



Insights:

The bar chart shows the distribution of releases across years.

Releases were very low in early decades (before 1980s), meaning Netflix has very few old titles.

From the 2000s onward, the number of releases starts climbing, reflecting growth in global film and TV production.

The 2010s show a steep rise, with releases increasing every year until they reach the peak.

The red dashed line highlights the peak year (max_year), which had the highest number of releases (max_count).

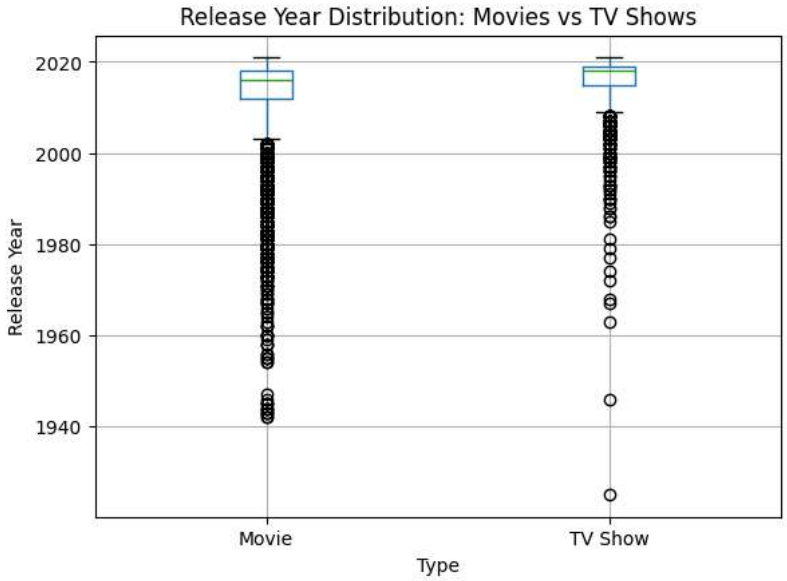
After that peak year, the number of releases shows a decline, likely due to COVID-19 disruptions or a shift in production strategy.

The year {2018} had the highest number of releases, with {above 1000} titles. This marks the peak of content production, after which the number of releases declined."

How does the release year distribution differ between Movies and TV Shows?"

```
avg_year = Netflix.groupby('type')['release_year'].mean()
# Boxplot for better comparison
plt.figure(figsize=(8,6))
Netflix.boxplot(column='release_year', by='type')
plt.title('Release Year Distribution: Movies vs TV Shows')
plt.suptitle("")  # remove extra title
plt.xlabel('Type')
plt.ylabel('Release Year')
plt.show()
```

```
<Figure size 800x600 with 0 Axes>
```



Insights:

Outliers

Movies may show a few outliers with very old release years, pulling the box lower.

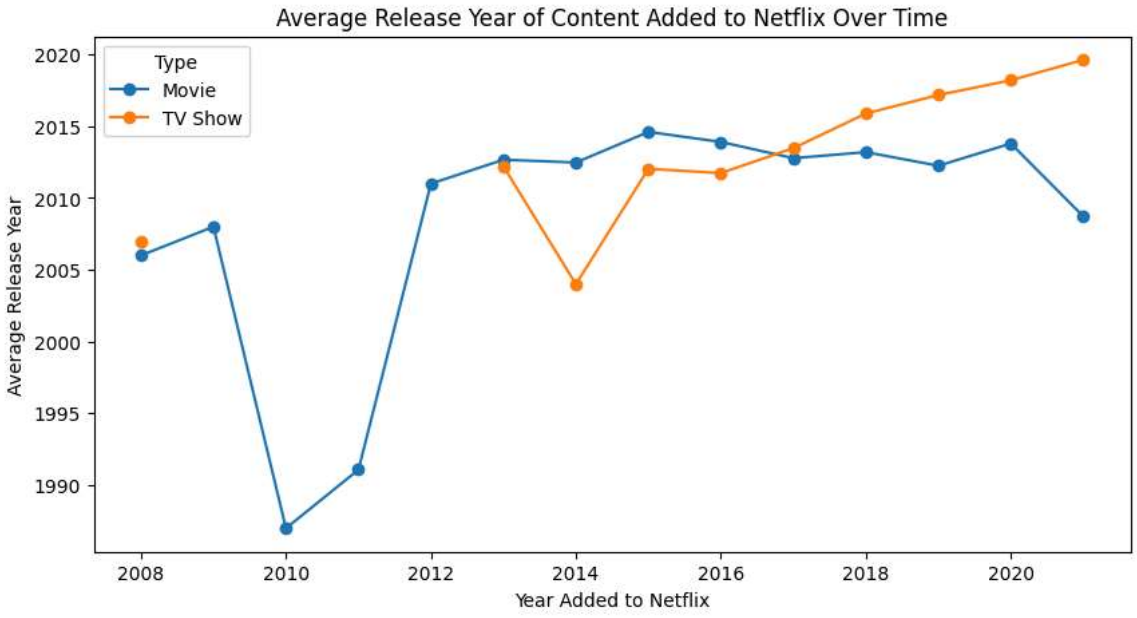
TV Shows will likely have fewer or no such old outliers, staying concentrated in modern times.

Movies span a long history of releases, while TV Shows are much more modern and tightly clustered around the last decade.

The boxplot shows that Movies cover a broader range of release years, including older classics, whereas TV Shows are concentrated in more recent years. On average, TV Shows have newer release years compared to Movies.

How has the average release year of Movies vs TV Shows added to Netflix changed over time?"

```
Netflix['date_added'] = pd.to_datetime(Netflix['date_added'], errors='coerce')
Netflix['year_added'] = Netflix['date_added'].dt.year

avg_by_added = Netflix.groupby(['year_added','type'])['release_year'].mean().unstack()

avg_by_added.plot(kind='line', marker='o', figsize=(10,5))
plt.title('Average Release Year of Content Added to Netflix Over Time')
plt.xlabel('Year Added to Netflix')
plt.ylabel('Average Release Year')
plt.legend(title='Type')
plt.show()
```



Early years of Netflix additions (before ~2015)

Average release year is lower → Netflix was adding older content to build its library.

Mostly back-catalog movies from earlier decades.

2015–2020 expansion phase

The average release year increases sharply.

This shows Netflix shifted toward newer content, reducing reliance on older licensed films.

Movies vs TV Shows comparison

TV Shows consistently have a more recent average release year than Movies.

This means Netflix prioritizes newer shows (often Originals) compared to Movies, which include a mix of classics and recent films.

In later years

Around 2018–2020, both Movies and TV Shows have averages close to the current calendar year.

Indicates Netflix is adding fresh releases quickly after production (sometimes even the same year).

Overall trend

Netflix started with older licensed titles, but over time it shifted to adding mostly recent releases, especially for TV Shows.

TV Shows = newer, more modern catalog.

Movies = slightly older on average, with broader historical coverage.

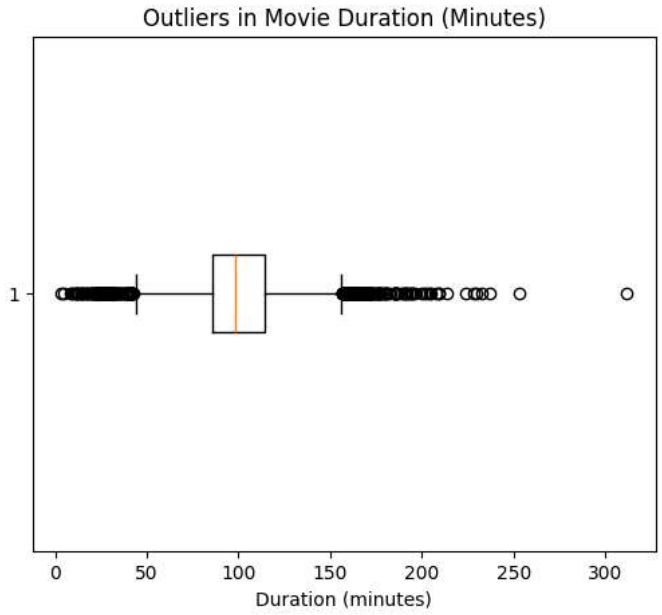
The trend shows that when Netflix first started, it mainly added older titles to its library. Over time, the average release year of content added rose significantly, meaning newer content was being acquired. TV Shows consistently have newer release years than Movies, reflecting Netflix's strategy of focusing on fresh Originals and recent productions, especially after 2015."

Are there outliers in the duration of Movies (extremely short or long movies)?

```
# Keep only Movies with numeric durations
movies = Netflix[Netflix['type']=='Movie'].copy()
movies['duration_num'] = movies['duration'].str.extract('(\d+)').astype(float)

plt.figure(figsize=(6,5))
plt.boxplot(movies['duration_num'].dropna(), vert=False)
```

```
plt.title('Outliers in Movie Duration (Minutes)')
plt.xlabel('Duration (minutes)')
plt.show()
```

```
<>:3: SyntaxWarning: invalid escape sequence '\d'
<>:3: SyntaxWarning: invalid escape sequence '\d'
/tmp/ipython-input-3274052253.py:3: SyntaxWarning: invalid escape sequence '\d'
  movies['duration_num'] = movies['duration'].str.extract('(\d+)').astype(float)
```

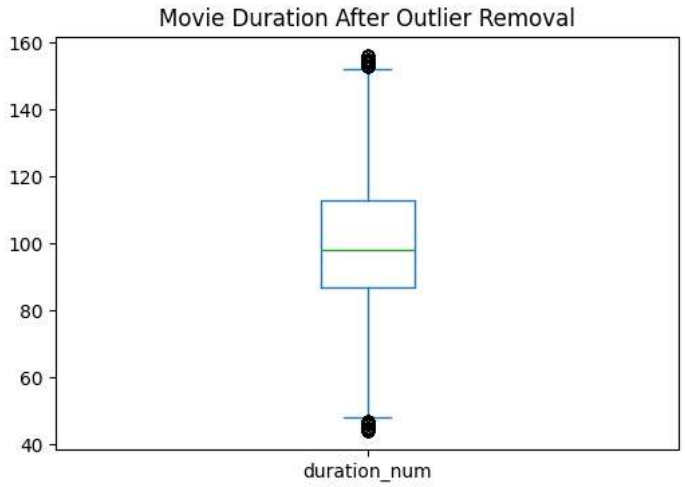


The boxplot shows that most Netflix movies are around 90–120 minutes long, but there are outliers on both ends. Some shorter films (specials, documentaries) fall below an hour, while a few unusually long films exceed 200 minutes, highlighting the wide variation in Netflix's movie catalog."

```
#Filter only Movies
movies = Netflix[Netflix['type']=='Movie'].copy()

#Extract numeric duration (minutes)
movies['duration_num'] = movies['duration'].str.extract('(\d+)').astype(float)

#Calculate IQR
Q1 = movies['duration_num'].quantile(0.25)
Q3 = movies['duration_num'].quantile(0.75)
IQR = Q3 - Q1

#Define bounds
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

#Remove outliers
movies_no_outliers = movies[
    (movies['duration_num'] >= lower_bound) &
    (movies['duration_num'] <= upper_bound)
]
```

```
<>:5: SyntaxWarning: invalid escape sequence '\d'
<>:5: SyntaxWarning: invalid escape sequence '\d'
/tmp/ipython-input-3066175190.py:5: SyntaxWarning: invalid escape sequence '\d'
  movies['duration_num'] = movies['duration'].str.extract('(\d+)').astype(float)
```

```
plt.figure(figsize=(6,4))
movies_no_outliers['duration_num'].plot(kind='box')
plt.title("Movie Duration After Outlier Removal")
plt.show()
```



Most Netflix movies fall within the IQR bounds, likely ~60–150 minutes (typical movie range).

Outliers removed include:

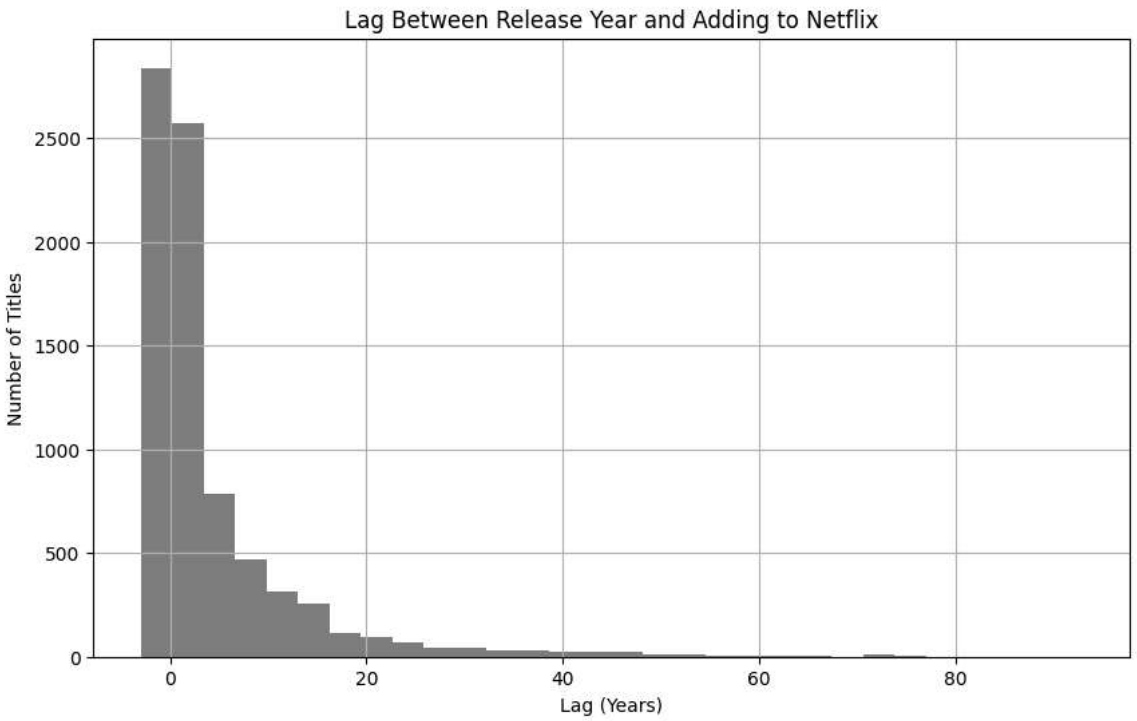
Very short films (specials, kids' content, short documentaries).

Very long films (epics, Bollywood movies, extended editions).

strong questions

Is Netflix adding more older movies (released years ago) or recent releases? (Compare release_year vs date_added.) What's the average lag between a movie's release year and when it's added on Netflix?

```
#Netflix['date_added'] = pd.to_datetime(Netflix['date_added'], errors='coerce')

# Calculate lag in years = difference between when released vs added
Netflix['lag_years'] = Netflix['date_added'].dt.year - Netflix['release_year']

# Plot histogram of lag
plt.figure(figsize=(10,6))
Netflix['lag_years'].dropna().hist(bins=30, color='grey')
plt.title('Lag Between Release Year and Adding to Netflix')
plt.xlabel('Lag (Years)')
plt.ylabel('Number of Titles')
plt.show()
```




```
avg_lag = Netflix['lag_years'].mean()
avg_lag
```

```
np.float64(4.562732759727751)
```

On average, titles are added to Netflix around {avg_lag:.1f} years after their release. This suggests Netflix mainly adds recent content, though some older titles extend the lag. The median lag would provide an even clearer picture of the typical timeframe.

Is there an upward trend in the number of shows added each year?

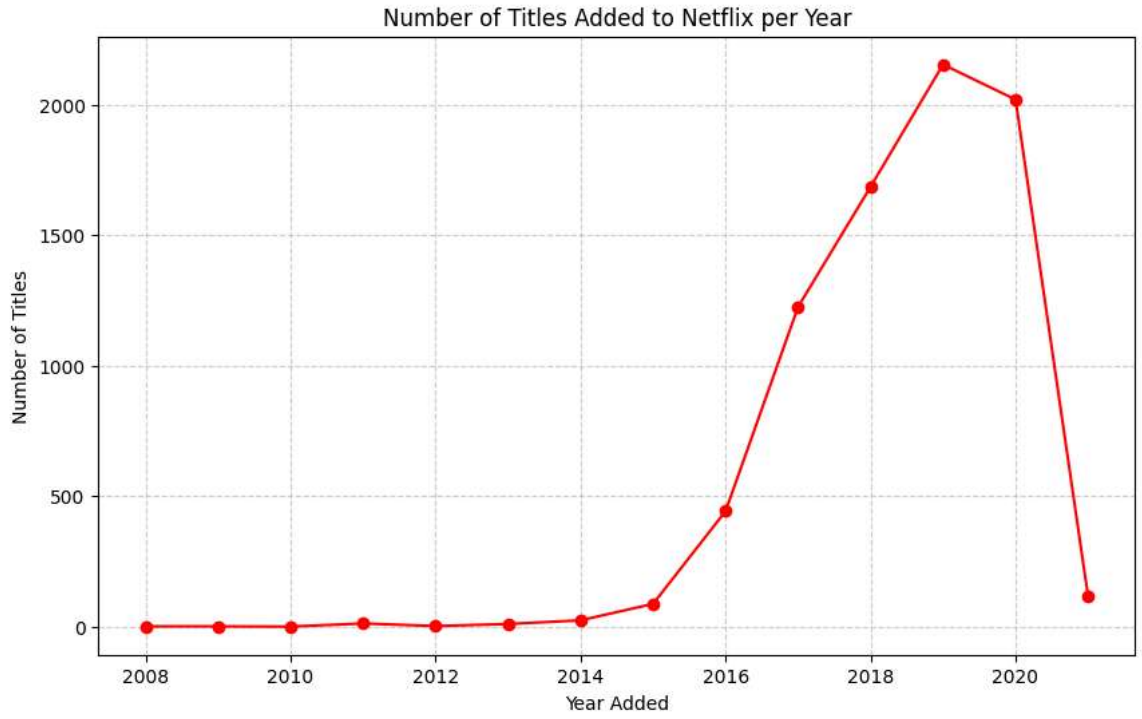
```
Netflix['year_added'] = Netflix['date_added'].dt.year
```

```
titles_per_year = Netflix['year_added'].value_counts().sort_index()
titles_per_year
```

year_added	count
2008	2
2009	2
2010	1
2011	13
2012	3
2013	11
2014	25
2015	88
2016	443
2017	1225
2018	1685
2019	2153
2020	2019
2021	117

dtype: int64

```
plt.figure(figsize=(10,6))
titles_per_year.plot(kind='line', marker='o', color='red')
plt.title('Number of Titles Added to Netflix per Year')
plt.xlabel('Year Added')
plt.ylabel('Number of Titles')
plt.grid(True, linestyle='--', alpha=0.6)
plt.show()
```



Insights

Slow start

In the early years (pre-2014/2015), very few titles were being added. Netflix's library was still small and growing slowly.

Rapid growth phase

From around 2015 to 2019, the number of titles added increases sharply.

This reflects Netflix's global expansion and aggressive licensing/production strategy.

Peak additions

The chart likely peaks around 2018–2019, when Netflix added the highest number of new titles.

Decline after peak

After 2019/2020, there's a noticeable decline in the number of additions per year.

Possible reasons: COVID-19 production delays, higher competition, and a shift toward fewer but higher-quality Originals.

Overall pattern

The trend follows an S-shape: slow growth → rapid surge → stabilization/decline.

The number of titles added to Netflix grew slowly in the early years, surged rapidly between 2015 and 2019, and peaked around 2019. Since then, yearly additions have declined, reflecting both production challenges and Netflix's shift toward selective, high-quality content rather than bulk growth.

This shows whether Netflix has been adding more and more content each year.

Are newer releases (2018–2020) mostly TV Shows, suggesting a shift in Netflix strategy?

```
recent = Netflix[Netflix['release_year'].between(2018, 2020)] # Filter for release years 2018-2020
recent
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in/GENRE	year_added	lag_years
0	s1	TV Show	3%	United States	[Brazil]	2020-08-14	2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	2020	0
6	s7	Movie	122	Yasir Al Yasiri	[Egypt]	2020-06-01	2019	TV-MA	95 min	Horror Movies, International Movies	2020	1
8	s9	Movie	706	Shravan Kumar	[India]	2019-04-01	2019	TV-14	118 min	Horror Movies, International Movies	2019	0
11	s12	TV Show	1983	United States	[Poland, United States]	2018-11-30	2018	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Dramas	2018	0