```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```python
df = pd.read_csv("netflix.csv")
```

```python
df
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...

Next steps: [Generate code with `df`] [⊙ View recommended plots]

```python
df.shape
```

```
(8807, 12)
```

```python
# exploding comma seperated values into rows and filling up null values for director column
df_director = df.loc[:,("show_id","director")]
df_director["director"] = df_director["director"].str.split(",")
df_director = df_director.explode("director")
df_director.head()
```

	show_id	director
0	s1	Kirsten Johnson
1	s2	NaN
2	s3	Julien Leclercq
3	s4	NaN
4	s5	NaN

Next steps: | Generate code with `df_director` | View recommended plots

```python
# exploding comma seperated values into rows and filling up null values for cast column
df_cast = df.loc[:,("show_id","cast")]
df_cast["cast"] = df_cast["cast"].str.split(",")
df_cast = df_cast.explode("cast")
df_cast.head()
```

	show_id	cast
0	s1	NaN
1	s2	Ama Qamata
1	s2	Khosi Ngema
1	s2	Gail Mabalane
1	s2	Thabang Molaba

Next steps: | Generate code with `df_cast` | View recommended plots

```python
# exploding comma seperated values into rows and filling up null values for country column
df_country = df.loc[:,("show_id","country")]
df_country["country"] = df_country["country"].str.split(",")
df_country = df_country.explode("country")
df_country.head()
```

	show_id	country
0	s1	United States
1	s2	South Africa
2	s3	NaN
3	s4	NaN
4	s5	India

Next steps: | Generate code with `df_country` | View recommended plots

```
# exploding comma seperated values into rows and filling up null values for listed_in column
df_listed_in = df.loc[:, ("show_id", "listed_in")]
df_listed_in["listed_in"] = df_listed_in["listed_in"].str.split(",")
df_listed_in = df_listed_in.explode("listed_in")
df_listed_in.head()
```

	show_id	listed_in
0	s1	Documentaries
1	s2	International TV Shows
1	s2	TV Dramas
1	s2	TV Mysteries
2	s3	Crime TV Shows

Next steps: Generate code with `df_listed_in`   View recommended plots

```
# merging 4 newly created dataframes in one using left join
df1 = pd.merge(df_director, df_cast, on = "show_id", how = "inner").merge(df_country, on = "show_id", how = "inner").merge(df_listed_in, on = "show_id", how = "inner")
df1.head()
```

	show_id	director	cast	country	listed_in
0	s1	Kirsten Johnson	NaN	United States	Documentaries
1	s2	NaN	Ama Qamata	South Africa	International TV Shows
2	s2	NaN	Ama Qamata	South Africa	TV Dramas
3	s2	NaN	Ama Qamata	South Africa	TV Mysteries
4	s2	NaN	Khosi Ngema	South Africa	International TV Shows

```
# removing columns from dataframes since they are explode[d]()
df2 = df.drop(columns = ["director", "cast", "country", "listed_in"])
df2.head()
```

	show_id	type	title	date_added	release_year	rating	duration	description
0	s1	Movie	Dick Johnson Is Dead	September 25, 2021	2020	PG-13	90 min	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	September 24, 2021	2021	TV-MA	1 Season	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	September 24, 2021	2021	TV-MA	1 Season	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	September 24, 2021	2021	TV-MA	2 Seasons	In a city of coaching centers known to train l...

Next steps: Generate code with `df2`   View recommended plots

```
# merging all the columns
netflix = df2.merge(df1, on = "show_id", how = "inner")
netflix
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in
0	s1	Movie	Dick Johnson Is Dead	September 25, 2021	2020	PG-13	90 min	As her father nears the end of his life, filmm...	Kirsten Johnson	NaN	United States	Documentaries
1	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Ama Qamata	South Africa	International TV Shows
2	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Ama Qamata	South Africa	TV Dramas
3	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Ama Qamata	South Africa	TV Mysteries
4	s2	TV Show	Blood & Water	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Khosi Ngema	South Africa	International TV Shows
...	...	...	...	...	...	...	...	...	...	...	...	...
202060	s8807	Movie	Zubaan	March 2, 2019	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Mozez Singh	Anita Shabdish	India	International Movies
202061	s8807	Movie	Zubaan	March 2, 2019	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Mozez Singh	Anita Shabdish	India	Music & Musicals
202062	s8807	Movie	Zubaan	March 2, 2019	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Mozez Singh	Chittaranjan Tripathy	India	Dramas

```
netflix.shape
```

```
(202065, 12)
```

```
netflix["date_added"] = pd.to_datetime(netflix["date_added"].str.strip(), format = "%B %d, %Y")
netflix.head()
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in
0	s1	Movie	Dick Johnson Is Dead	2021-09-25	2020	PG-13	90 min	As her father nears the end of his life, filmm...	Kirsten Johnson	NaN	United States	Documentaries
1	s2	TV Show	Blood & Water	2021-09-24	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Ama Qamata	South Africa	International TV Shows
2	s2	TV Show	Blood & Water	2021-09-24	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Ama Qamata	South Africa	TV Dramas
3	s2	TV Show	Blood & Water	2021-09-24	2021	TV-MA	2 Seasons	After crossing paths at a...	NaN	Ama	South	TV Mysteries

```
netflix["release_year"] = pd.to_datetime(netflix["release_year"], format = "%Y").dt.year
netflix.head()
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in	
0	s1	Movie	Dick Johnson Is Dead	2021-09-25	2020	PG-13	90 min	As her father nears the end of his life, filmm...	Kirsten Johnson	NaN	United States	Documentaries	
1	s2	TV Show	Blood & Water	2021-09-24	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Ama Qamata	South Africa	International TV Shows	
2	s2	TV Show	Blood & Water	2021-09-24	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Ama Qamata	South Africa	TV Dramas	
	s2	TV	Blood &	2021-09-24	2021	TV-MA	2	After crossing paths at a	NaN	Ama	South	TV M	

```
netflix.isna().sum()
```

```
show_id            0
type               0
title              0
date_added       158
release_year       0
rating            67
duration           3
description        0
director       50643
cast            2149
country        11897
listed_in          0
dtype: int64
```

```
# there are some null values in rating we can fill these values by mode of ratings of movies produced by certain director or cast
netflix[netflix["rating"].isna()]
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in
135172	s5990	Movie	13TH: A Conversation with Oprah Winfrey & Ava ...	2017-01-26	2017	NaN	37 min	Oprah Winfrey sits down with director Ava DuVe...	NaN	Oprah Winfrey	NaN	Movies
135173	s5990	Movie	13TH: A Conversation with Oprah Winfrey & Ava ...	2017-01-26	2017	NaN	37 min	Oprah Winfrey sits down with director Ava DuVe...	NaN	Ava DuVernay	NaN	Movies
154424	s6828	TV Show	Gargantia on the Verdurous Planet	2016-12-01	2013	NaN	1 Season	After falling through a wormhole, a space-dwel...	NaN	Kaito Ishikawa	Japan	Anime Series
154425	s6828	TV Show	Gargantia on the Verdurous Planet	2016-12-01	2013	NaN	1 Season	After falling through a wormhole, a space-dwel...	NaN	Kaito Ishikawa	Japan	International TV Shows
154426	s6828	TV Show	Gargantia on the Verdurous Planet	2016-12-01	2013	NaN	1 Season	After falling through a wormhole, a space-dwel...	NaN	Hisako Kanemoto	Japan	Anime Series
...	...	...	...	...	...	...	...	...	...	...	...	...
172016	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	NaN	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Francesco Migliore	Italy	Dramas
172017	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	NaN	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Albrecht Weimer	Italy	Dramas
172018	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	NaN	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Giulia	Italy	D...

```
# replacing null values of rating of movies directed by Alessandro Pepe
netflix["director"][netflix["rating"].isna()].value_counts()
```

```
director
Alessandro Pepe    7
Name: count, dtype: int64
```

```
netflix[netflix["director"] == "Alessandro Pepe"]
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in
172014	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	NaN	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Leone Frisa	Italy	Dramas
172015	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	NaN	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Paolo Vaccarino	Italy	Dramas
172016	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	NaN	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Francesco Migliore	Italy	Dramas
172017	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	NaN	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Albrecht Weimer	Italy	Dramas
172018	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	NaN	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Giulia Dichiaro	Italy	Dramas

Since movie produced by Alessandro Pepe is listed_in Dramas catergory find out the mode of ratings for this "Drama" category

```
netflix["rating"][netflix["listed_in"] == "Dramas"].value_counts()
```

```
    rating
    TV-MA    7465
    TV-14    4208
    R        3988
    PG-13    2097
    TV-PG    1048
    PG        410
    NR        265
    TV-G       98
    NC-17      39
    UR         22
    TV-Y7      10
    Name: count, dtype: int64
```

Mode of Drama Category is TV-MA or TV - 14. Reading at at the movie description TV - MA is more suitable rating.

```
netflix.loc[netflix['director'] == 'Alessandro Pepe'] = netflix.loc[netflix['director'] == 'Alessandro Pepe'] .fillna("TV-MA")
```

```
netflix.loc[netflix['director'] == 'Alessandro Pepe']
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in
172014	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	TV-MA	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Leone Frisa	Italy	Dramas
172015	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	TV-MA	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Paolo Vaccarino	Italy	Dramas
172016	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	TV-MA	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Francesco Migliore	Italy	Dramas
172017	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	TV-MA	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Albrecht Weimer	Italy	Dramas
172018	s7538	Movie	My Honor Was Loyalty	2017-03-01	2015	TV-MA	115 min	Amid the chaos and horror of World War II, a c...	Alessandro Pepe	Giulia Dichiaro	Italy	Dramas

```
netflix["rating"][netflix["listed_in"] == "Anime Series"].value_counts()
```

```
    rating
    TV-MA    899
    TV-14    871
    TV-Y7    298
    TV-PG    210
    TV-Y      13
    Name: count, dtype: int64
```

```
netflix[netflix["rating"].isna()]
netflix.loc[netflix['show_id'] == 's6828'] = netflix.loc[netflix['show_id'] == 's6828'].fillna("TV-MA")
```

```python
netflix["rating"][netflix["listed_in"] == "Kids' TV"].value_counts()
```

```
        rating
TV-Y7     2005
TV-Y      1081
TV-PG      378
TV-G       330
TV-Y7-FV     8
Name: count, dtype: int64
```

```python
netflix.loc[netflix['show_id'] == 's7313'] = netflix.loc[netflix['show_id'] == 's7313'].fillna("TV-Y7")
```

```python
netflix[netflix["cast"] == "Oprah Winfrey"]
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in
89508	s3757	Movie	Oprah Presents When They See Us Now	2019-06-13	2019	TV-PG	61 min	Oprah Winfrey talks with the exonerated men on...	Mark Ritchie	Oprah Winfrey	United States	Documentaries
135172	s5990	Movie	13TH: A Conversation with Oprah Winfrey &	2017-01-26	2017	NaN	37 min	Oprah Winfrey sits down with director	NaN	Oprah Winfrey	NaN	Movies

```python
netflix.loc[netflix['show_id'] == 's5990'] = netflix.loc[netflix['show_id'] == 's5990'].fillna("TV-PG")
```

```python
netflix.isna().sum()
```

```
show_id          0
type             0
title            0
date_added     158
release_year     0
rating           0
duration         3
description      0
director     50583
cast          2149
country      11895
listed_in        0
dtype: int64
```

```python
netflix[netflix["duration"].isna()]
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in
126582	s5542	Movie	Louis C.K. 2017	2017-04-04	2017	74 min	NaN	Louis C.K. muses on religion, eternal love, gi...	Louis C.K.	Louis C.K.	United States	Movies
131648	s5795	Movie	Louis C.K.: Hilarious	2016-09-16	2010	84 min	NaN	Emmy-winning comedy writer Louis C.K. brings h...	Louis C.K.	Louis C.K.	United States	Movies

```python
netflix.loc[netflix["director"] ==  "Louis C.K."]
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in
126582	s5542	Movie	Louis C.K. 2017	2017-04-04	2017	74 min	NaN	Louis C.K. muses on religion, eternal love, gi...	Louis C.K.	Louis C.K.	United States	Movies
131648	s5795	Movie	Louis C.K.: Hilarious	2016-09-16	2010	84 min	NaN	Emmy-winning comedy writer Louis C.K. brings h...	Louis C.K.	Louis C.K.	United States	Movies

Rating column should be in duration and rating should be null or fillna("TV-MA") (since most of content is TV-MA)

```
netflix.loc[netflix["director"] == "Louis C.K.", "duration"] = netflix.loc[netflix["director"] ==  "Louis C.K.", "rating"]
netflix.loc[netflix["director"] == "Louis C.K.", "rating"] = "TV-MA"
```

```
netflix.loc[netflix["director"] == "Louis C.K."]
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in
126582	s5542	Movie	Louis C.K. 2017	2017-04-04	2017	TV-MA	74 min	Louis C.K. muses on religion, eternal love, gi...	Louis C.K.	Louis C.K.	United States	Movies
131648	s5795	Movie	Louis C.K.: Hilarious	2016-09-16	2010	TV-MA	84 min	Emmy-winning comedy writer Louis C.K. brings h...	Louis C.K.	Louis C.K.	United States	Movies

```
netflix.isna().sum()
```

```
show_id          0
type             0
title            0
date_added     158
release_year     0
rating           0
duration         0
description      0
director     50583
cast          2149
country      11895
listed_in        0
dtype: int64
```

```
netflix.head()
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in
0	s1	Movie	Dick Johnson Is Dead	2021-09-25	2020	PG-13	90 min	As her father nears the end of his life, filmm...	Kirsten Johnson	NaN	United States	Documentaries
1	s2	TV Show	Blood & Water	2021-09-24	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Ama Qamata	South Africa	International TV Shows
2	s2	TV Show	Blood & Water	2021-09-24	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Ama Qamata	South Africa	TV Dramas
3	s	TV	Blood &	2021-09-24	2021	TV-MA	2	After crossing paths at a	NaN	Ama	South	TV Mysteries

```
# Stripping the Extra spaces in data
netflix.loc[:, "director"] = netflix.loc[:, "director"].str.strip()
netflix.loc[:, "cast"] = netflix.loc[:, "cast"].str.strip()
netflix.loc[:, "country"] = netflix.loc[:, "country"].str.strip()
netflix.loc[:, "listed_in"] = netflix.loc[:, "listed_in"].str.strip()
netflix.loc[:, "title"] = netflix.loc[:, "title"].str.strip()
```

```
netflix.head()
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in
0	s1	Movie	Dick Johnson Is Dead	2021-09-25	2020	PG-13	90 min	As her father nears the end of his life, filmm...	Kirsten Johnson	NaN	United States	Documentaries
1	s2	TV Show	Blood & Water	2021-09-24	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Ama Qamata	South Africa	International TV Shows
2	s2	TV Show	Blood & Water	2021-09-24	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	NaN	Ama Qamata	South Africa	TV Dramas
3	s	TV	Blood &	2021-09-24	2021	TV-MA	2	After crossing paths at a	NaN	Ama	South	TV Dramas

```
netflix.isna().sum()
```

```
show_id             0
type                0
title               0
date_added        158
release_year        0
rating              0
duration            0
description         0
director        50583
cast             2149
country         11895
listed_in           0
dtype: int64
```

```
# fill null values by mode of country in which director has produced movies director name
def mode(x):
  return x.mode()
cf = netflix.groupby('director')['country'].apply(mode).reset_index()
for director in netflix[netflix["country"].isna()]["director"]:
  if director in cf["director"].values:
    netflix["country"].fillna(cf[cf["director"] == director]["country"].values[0], inplace = True)
```

```
netflix.isna().sum()
```

```
show_id             0
type                0
title               0
date_added        158
release_year        0
rating              0
duration            0
description         0
director        50583
cast             2149
```

```
        country          0
        listed_in        0
        dtype: int64
```

```python
# separating columns for year, month, week of the year using date_added column
netflix["date_added_year"] = netflix["date_added"].dt.year
netflix["date_added_month"] = netflix["date_added"].dt.month
netflix["date_added_week"] = netflix["date_added"].dt.isocalendar().week
```

```python
# filling null values in director and cast column.
netflix["director"].fillna("Unknown director", inplace = True)
netflix["cast"].fillna("Unknown cast", inplace = True)
```

```python
# filling null values in date_added column with release year.
netflix.loc[netflix["date_added"].isna(), "date_added"] = netflix.loc[netflix["date_added"].isna(), "release_year"]
netflix.loc[netflix["date_added_year"].isna(), "date_added_year"] = netflix.loc[netflix["date_added_year"].isna(), "release_year"]
```

```python
# filling month and week column by 1-1
netflix["date_added_month"].fillna(1.0, inplace = True)
netflix["date_added_week"].fillna(1, inplace = True)
```

```python
netflix.isna().sum()
```

```
        show_id          0
        type             0
        title            0
        date_added       0
        release_year     0
        rating           0
        duration         0
        description      0
        director         0
        cast             0
        country          0
        listed_in        0
        date_added_year  0
        date_added_month 0
        date_added_week  0
        dtype: int64
```

```python
# Splitting the dataframe into 2 type 1st Movies, 2nd TV Show
df_movies = netflix.loc[netflix["type"] == "Movie"].reset_index()
```

```python
df_shows = netflix.loc[netflix["type"] == "TV Show"].reset_index()
netflix = pd.concat([df_movies, df_shows], axis=0).reset_index()
```

```python
df_movies
```

	index	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in	date_added_y
0	0	s1	Movie	Dick Johnson Is Dead	2021-09-25 00:00:00	2020	PG-13	90 min	As her father nears the end of his life, filmm...	Kirsten Johnson	Unknown cast	United States	Documentaries	20;
1	159	s7	Movie	My Little Pony: A New Generation	2021-09-24 00:00:00	2021	PG	91 min	Equestria's divided. But a bright-eyed hero be...	Robert Cullen	Vanessa Hudgens	France	Children & Family Movies	20;
2	160	s7	Movie	My Little Pony: A New Generation	2021-09-24 00:00:00	2021	PG	91 min	Equestria's divided. But a bright-eyed hero be...	Robert Cullen	Kimiko Glenn	France	Children & Family Movies	20;
3	161	s7	Movie	My Little Pony: A New Generation	2021-09-24 00:00:00	2021	PG	91 min	Equestria's divided. But a bright-eyed hero be...	Robert Cullen	James Marsden	France	Children & Family Movies	20;
4	162	s7	Movie	My Little Pony: A New Generation	2021-09-24 00:00:00	2021	PG	91 min	Equestria's divided. But a bright-eyed hero be...	Robert Cullen	Sofia Carson	France	Children & Family Movies	20;
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
145912	202060	s8807	Movie	Zubaan	2019-03-02 00:00:00	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Mozez Singh	Anita Shabdish	India	International Movies	20
145913	202061	s8807	Movie	Zubaan	2019-03-02 00:00:00	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Mozez Singh	Anita Shabdish	India	Music & Musicals	20
145914	202062	s8807	Movie	Zubaan	2019-03-02 00:00:00	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Mozez Singh	Chittaranjan Tripathy	India	Dramas	20
145915	202063	s8807	Movie	Zubaan	2019-03-02 00:00:00	2015	TV-14	111 min	A scrappy but poor boy worms his way into a ty...	Mozez Singh	Chittaranjan Tripathy	India	International Movies	20

df_shows

	index	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in	date_added_year	date
0	1	s2	TV Show	Blood & Water	2021-09-24 00:00:00	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Unknown director	Ama Qamata	South Africa	International TV Shows	2021.0	
1	2	s2	TV Show	Blood & Water	2021-09-24 00:00:00	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Unknown director	Ama Qamata	South Africa	TV Dramas	2021.0	
2	3	s2	TV Show	Blood & Water	2021-09-24 00:00:00	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Unknown director	Ama Qamata	South Africa	TV Mysteries	2021.0	
3	4	s2	TV Show	Blood & Water	2021-09-24 00:00:00	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Unknown director	Khosi Ngema	South Africa	International TV Shows	2021.0	
4	5	s2	TV Show	Blood & Water	2021-09-24 00:00:00	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	Unknown director	Khosi Ngema	South Africa	TV Dramas	2021.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
56143	201938	s8801	TV Show	Zindagi Gulzar Hai	2016-12-15 00:00:00	2012	TV-PG	1 Season	Strong-willed, middle-class Kashaf and carefre...	Unknown director	Hina Khawaja Bayat	Pakistan	Romantic TV Shows	2016.0	
56144	201939	s8801	TV Show	Zindagi Gulzar Hai	2016-12-15 00:00:00	2012	TV-PG	1 Season	Strong-willed, middle-class Kashaf and carefre...	Unknown director	Hina Khawaja Bayat	Pakistan	TV Dramas	2016.0	

Next steps: Generate code with `df_shows`   ⬤ View recommended plots

```
netflix.drop(columns = ["level_0", "index"], inplace = True)

netflix.head()
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in	date_added_year	date_added
0	s1	Movie	Dick Johnson Is Dead	2021-09-25 00:00:00	2020	PG-13	90 min	As her father nears the end of his life, filmm...	Kirsten Johnson	Unknown cast	United States	Documentaries	2021.0	
1	s7	Movie	My Little Pony: A New Generation	2021-09-24 00:00:00	2021	PG	91 min	Equestria's divided. But a bright-eyed hero be...	Robert Cullen	Vanessa Hudgens	France	Children & Family Movies	2021.0	
2	s7	Movie	My Little Pony: A New Generation	2021-09-24 00:00:00	2021	PG	91 min	Equestria's divided. But a bright-eyed hero be...	Robert Cullen	Kimiko Glenn	France	Children & Family Movies	2021.0	
3	s7	Movie	My Little Pony: A New	2021-09-24 00:00:00	2021	PG	91 min	Equestria's divided. But a bright-eyed	Robert Cullen	James Marsden	France	Children & Family Movies	2021.0	

```
# Removing duplicate rows in the dataset
netflix = netflix.drop_duplicates(keep = "last")
```

```
netflix.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 202010 entries, 0 to 202064
Data columns (total 15 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   show_id           202010 non-null  object
 1   type              202010 non-null  object
 2   title             202010 non-null  object
 3   date_added        202010 non-null  object
 4   release_year      202010 non-null  int32
 5   rating            202010 non-null  object
 6   duration          202010 non-null  object
 7   description       202010 non-null  object
 8   director          202010 non-null  object
 9   cast              202010 non-null  object
 10  country           202010 non-null  object
 11  listed_in         202010 non-null  object
 12  date_added_year   202010 non-null  float64
 13  date_added_month  202010 non-null  float64
 14  date_added_week   202010 non-null  UInt32
dtypes: UInt32(1), float64(2), int32(1), object(11)
memory usage: 23.3+ MB
```
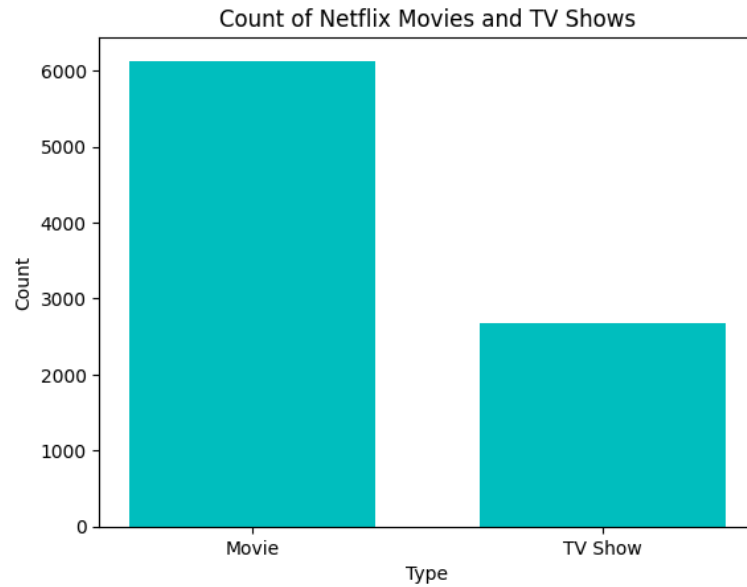
```
# Shows/ Movies Analysis
# What are counts of Movies vs different TV shows
netflix.groupby('type')['show_id'].nunique()
```

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

```
plt.bar(netflix.groupby('type')['show_id'].nunique().index, netflix.groupby('type')['show_id'].nunique().values, color = ["c"], width = 0.7)
plt.title('Count of Netflix Movies and TV Shows')
plt.xlabel('Type')
plt.ylabel('Count')

plt.show()
```
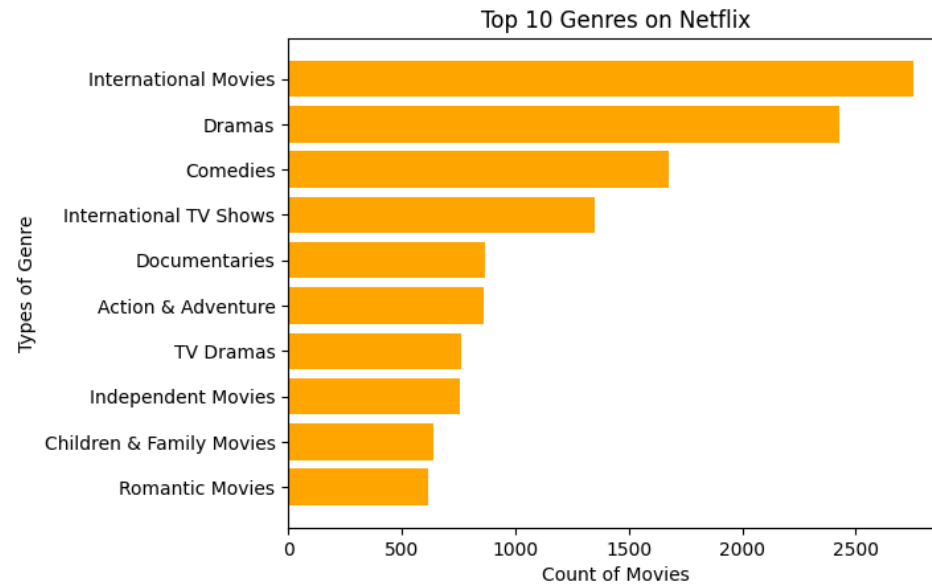


```
# Counts of movies/tv shows by each Genre
barh = netflix.groupby(['listed_in']).agg({"title":"nunique"}).sort_values(by = "title", ascending = False).reset_index()
barh.head(10)
```

	listed_in	title
0	International Movies	2751
1	Dramas	2426
2	Comedies	1674
3	International TV Shows	1351
4	Documentaries	869
5	Action & Adventure	859
6	TV Dramas	763
7	Independent Movies	756
8	Children & Family Movies	641
9	Romantic Movies	616

Next steps: Generate code with `barh`   ◉ View recommended plots

```
# Top 10 Genres on Netflix
plt.barh(barh.head(10)['listed_in'][::-1], barh.head(10)["title"][::-1], color = ["orange"])
plt.title('Top 10 Genres on Netflix')
plt.ylabel('Types of Genre')
plt.xlabel('Count of Movies')
plt.show()
```


Top 10 Genres on Netflix
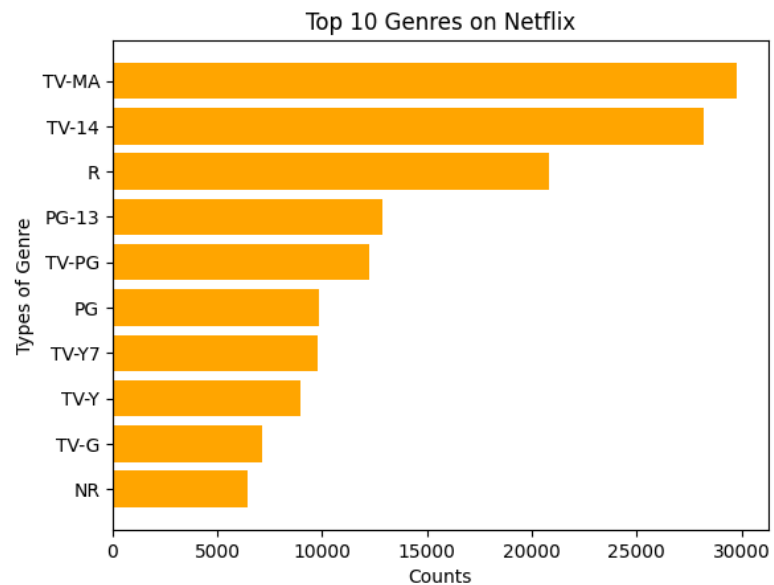
```
# Counts of rating of Movies/TV shows
netflix["rating"].value_counts()
```

```
        rating
        TV-MA       73921
        TV-14       43951
        R           25859
        PG-13       16246
        TV-PG       14928
        PG          10919
        TV-Y7        6318
        TV-Y         3665
        TV-G         2779
        NR           1573
        G            1530
        NC-17         149
        TV-Y7-FV       86
        UR             86
        Name: count, dtype: int64
```

```
# Top 10 Ratings on Netflix
plt.barh(netflix["rating"].value_counts()[:10].index[::-1], netflix["listed_in"].value_counts()[:10].values[::-1], color = ["orange"])
plt.title('Top 10 Genres on Netflix')
plt.xlabel('Counts')
plt.ylabel('Types of Genre')
plt.show()
```



```
netflix.head()
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in	date_added_year	date_added
0	s1	Movie	Dick Johnson Is Dead	2021-09-25 00:00:00	2020	PG-13	90 min	As her father nears the end of his life, filmm...	Kirsten Johnson	Unknown cast	United States	Documentaries	2021.0	
1	s7	Movie	My Little Pony: A New Generation	2021-09-24 00:00:00	2021	PG	91 min	Equestria's divided. But a bright-eyed hero be...	Robert Cullen	Vanessa Hudgens	France	Children & Family Movies	2021.0	
2	s7	Movie	My Little Pony: A New Generation	2021-09-24 00:00:00	2021	PG	91 min	Equestria's divided. But a bright-eyed hero be...	Robert Cullen	Kimiko Glenn	France	Children & Family Movies	2021.0	
3	s7	Movie	My Little Pony: A New	2021-09-24 00:00:00	2021	PG	91 min	Equestria's divided. But	Robert Cullen	James Marsden	France	Children & Family Movies	2021.0	

Duration column has minutes and seasons written, we need to remove there strings and convert it to integers

```
netflix["duration"] = netflix["duration"].str.split(" ", expand = True).loc[:, 0].astype("int")
df_movies["duration"] = df_movies["duration"].str.split(" ", expand = True).loc[:, 0].astype("int")
df_shows["duration"] = df_shows["duration"].str.split(" ", expand = True).loc[:, 0].astype("int")
df_movies.drop(columns = ["index"])
df_shows.drop(columns = ["index"])
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in	date_added_year	date_added_m
0	s2	TV Show	Blood & Water	2021-09-24 00:00:00	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...	Unknown director	Ama Qamata	South Africa	International TV Shows	2021.0	
1	s2	TV Show	Blood & Water	2021-09-24 00:00:00	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...	Unknown director	Ama Qamata	South Africa	TV Dramas	2021.0	
2	s2	TV Show	Blood & Water	2021-09-24 00:00:00	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...	Unknown director	Ama Qamata	South Africa	TV Mysteries	2021.0	
3	s2	TV Show	Blood & Water	2021-09-24 00:00:00	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...	Unknown director	Khosi Ngema	South Africa	International TV Shows	2021.0	
4	s2	TV Show	Blood & Water	2021-09-24 00:00:00	2021	TV-MA	2	After crossing paths at a party, a Cape Town t...	Unknown director	Khosi Ngema	South Africa	TV Dramas	2021.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
56143	s8801	TV Show	Zindagi Gulzar Hai	2016-12-15 00:00:00	2012	TV-PG	1	Strong-willed, middle-class Kashaf and carefre...	Unknown director	Hina Khawaja Bayat	Pakistan	Romantic TV Shows	2016.0	

Since I have seperated movies and TV shows columns we can analyse them seperatly

```
pie = netflix.groupby(['type']).agg({"title":"nunique"}).reset_index()
pie
```
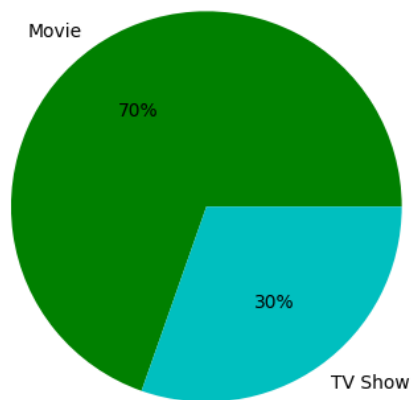
	type	title
0	Movie	6130
1	TV Show	2676

Next steps: [Generate code with `pie`]   [◉ View recommended plots]

```python
plt.pie(pie['title'], labels = pie["type"],autopct='%.0f%%', colors=['green','c'])
plt.show()
```


We have approx 7 : 3 ratio of Movies to the TV Shows in Netflix

```python
netflix.dtypes
```

```
show_id              object
type                 object
title                object
date_added           object
release_year          int32
rating               object
duration              int64
description          object
director             object
cast                 object
country              object
listed_in            object
date_added_year     float64
date_added_month    float64
date_added_week      UInt32
dtype: object
```
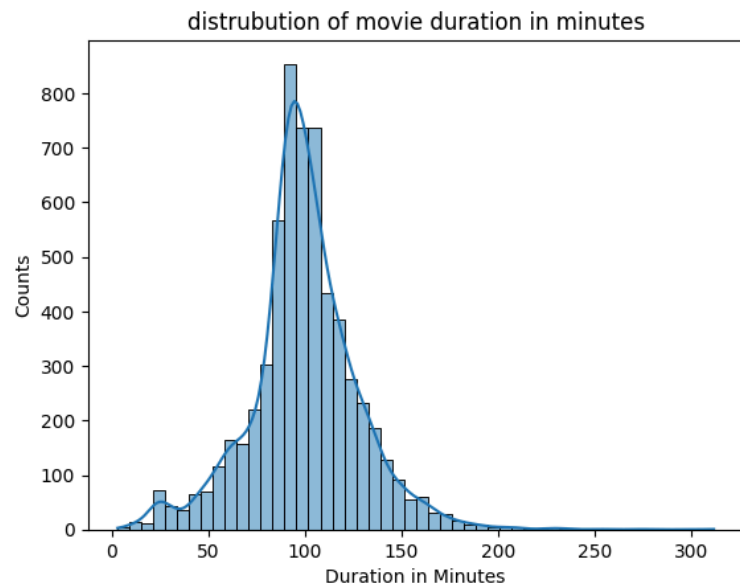
```python
df_movies[["show_id","duration"]].groupby("show_id").min().describe()
```

	duration
count	6131.000000
mean	99.564998
std	28.289504
min	3.000000
25%	87.000000
50%	98.000000
75%	114.000000
max	312.000000

Average duration on movies on netflix is 99.5 mins, with standard deviation of 28. 2 mins.

```
import seaborn as sns
sns.histplot(df_movies[["show_id","duration"]].groupby("show_id").min().reset_index().loc[:, ("duration")],  kde = True, bins = 50)
plt.title('distrubution of movie duration in minutes')
plt.ylabel('Counts')
plt.xlabel('Duration in Minutes')
plt.show()
```



```
df_shows[["show_id","duration"]].groupby("show_id").max().describe()
```

	duration
count	2676.000000
mean	1.764948
std	1.582752
min	1.000000
25%	1.000000
50%	1.000000
75%	2.000000
max	17.000000

There is mean of 1.7 seasons per TV-show.

```
netflix.head()
```

	show_id	type	title	date_added	release_year	rating	duration	description	director	cast	country	listed_in	date_added_year	date_added_
0	s1	Movie	Dick Johnson Is Dead	2021-09-25 00:00:00	2020	PG-13	90	As her father nears the end of his life, filmm...	Kirsten Johnson	Unknown cast	United States	Documentaries	2021.0	
1	s7	Movie	My Little Pony: A New Generation	2021-09-24 00:00:00	2021	PG	91	Equestria's divided. But a bright-eyed hero be...	Robert Cullen	Vanessa Hudgens	France	Children & Family Movies	2021.0	
2	s7	Movie	My Little Pony: A New Generation	2021-09-24 00:00:00	2021	PG	91	Equestria's divided. But a bright-eyed hero be...	Robert Cullen	Kimiko Glenn	France	Children & Family Movies	2021.0	
3	s7	Movie	My Little Pony: A New Generation	2021-09-24 00:00:00	2021	PG	91	Equestria's divided. But a bright-eyed hero be...	Robert Cullen	James Marsden	France	Children & Family Movies	2021.0	

```
# number of movies produced by each countries.
country = df_movies.groupby("country").agg({"show_id":"nunique"}).reset_index().sort_values(by = "show_id", ascending = False)
country
```
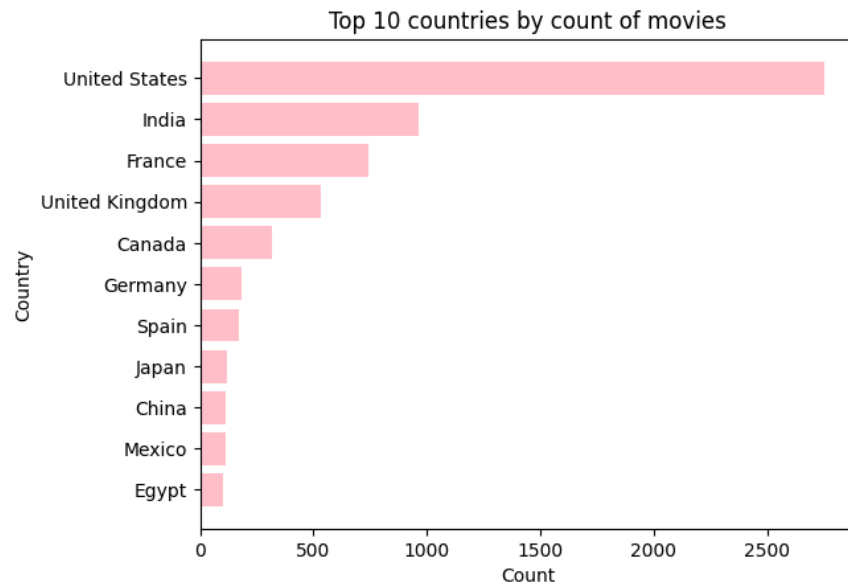
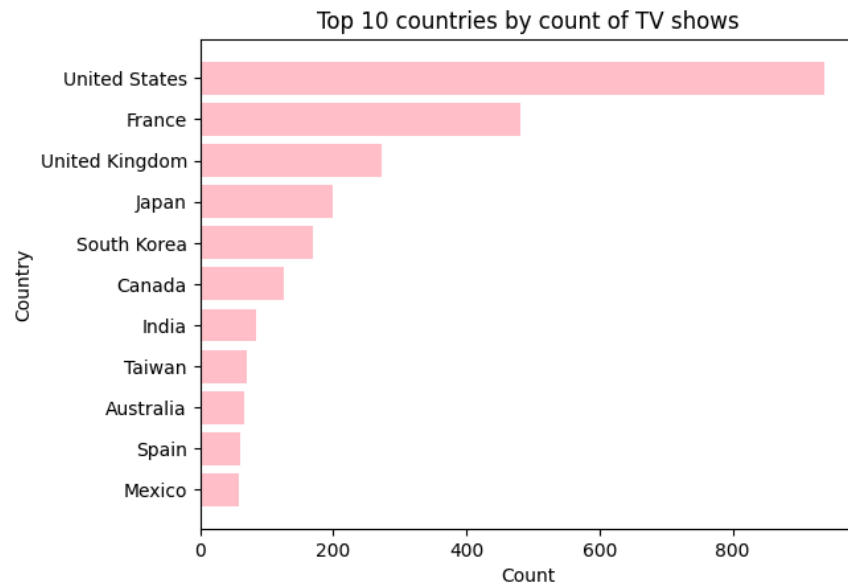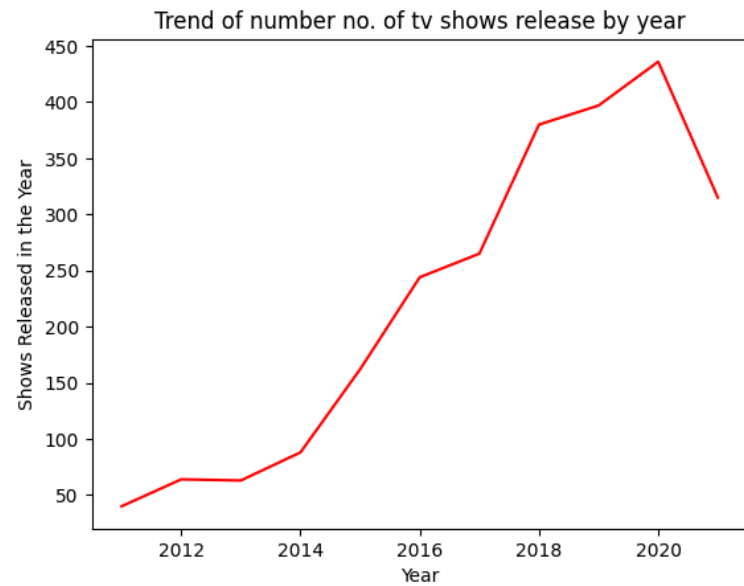|	country	show_id
112	United States	2752
42	India	962
33	France	742
111	United Kingdom	534
19	Canada	319
...	...	...
31	Ethiopia	1
87	Samoa	1
29	Ecuador	1
27	Dominican Republic	1
28	East Germany	1

119 rows × 2 columns

Next steps: Generate code with `country`   View recommended plots
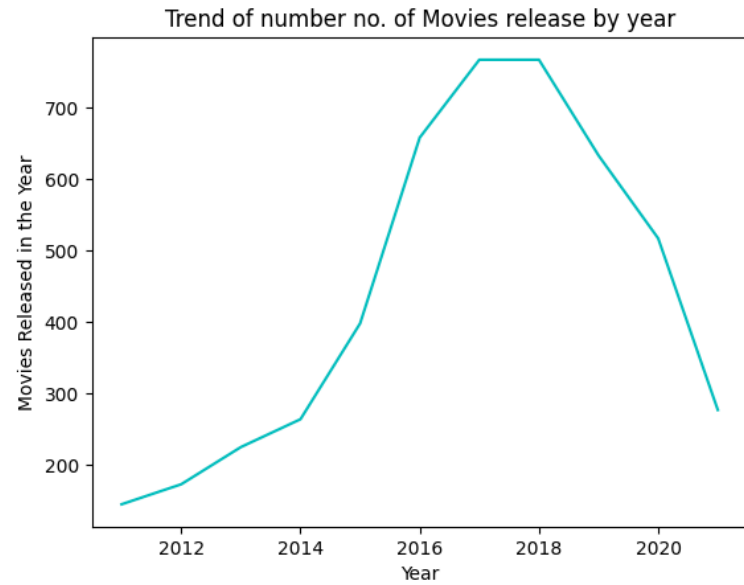
```
plt.barh(country["country"][:11][::-1], country["show_id"][:11][::-1], color = "pink")
plt.title("Top 10 countries by count of movies")
plt.xlabel("Count")
plt.ylabel("Country")
plt.show()
```


Top 10 countries by count of movies

```
countru_shows = df_shows.groupby("country").agg({"show_id":"nunique"}).reset_index().sort_values(by = "show_id", ascending = False)
countru_shows
```

	country	show_id
63	United States	938
19	France	481
62	United Kingdom	272
30	Japan	199
52	South Korea	170
...	...	...
55	Switzerland	1
23	Hungary	1
36	Malta	1
49	Senegal	1
0		1

66 rows × 2 columns

Next steps: Generate code with `countru_shows`   View recommended plots

```
plt.barh(countru_shows["country"][:11][::-1], countru_shows["show_id"][:11][::-1], color = "pink")
plt.title("Top 10 countries by count of TV shows")
plt.xlabel("Count")
plt.ylabel("Country")
plt.show()
```
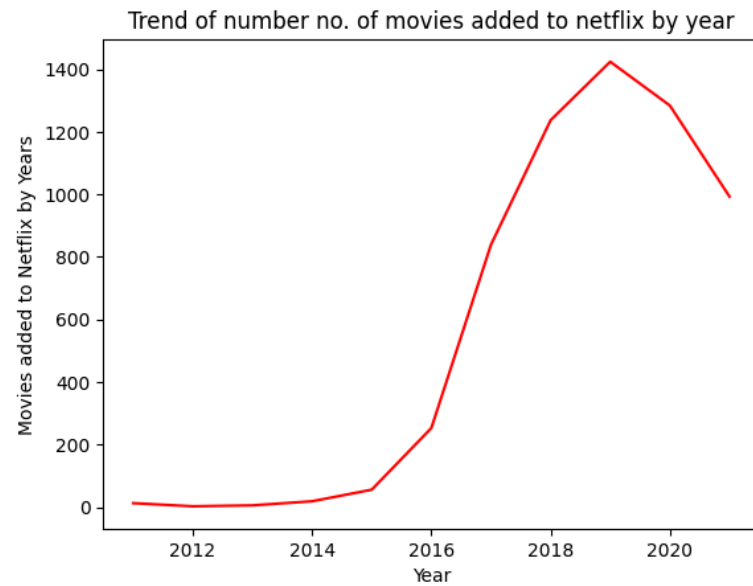
Top 10 countries by count of TV shows


```
#  trend of number no tv show release by year
release_year = df_shows.groupby("release_year").agg({"show_id": "nunique"})[df_shows.groupby("release_year").agg({"show_id": "nunique"}).index > 2010]
sns.lineplot(data=release_year, x='release_year', y='show_id', color = "r")
plt.ylabel("Shows Released in the Year")
plt.xlabel("Year")
plt.title("Trend of number no. of tv shows release by year")
plt.show()
```

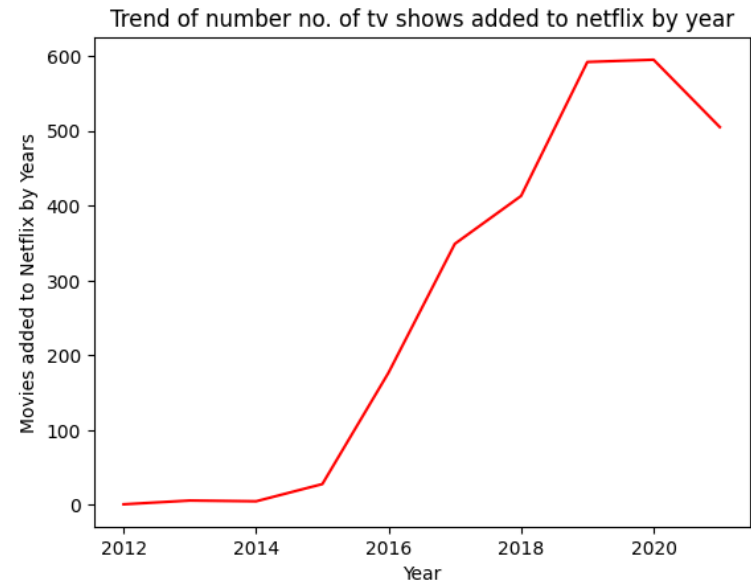Trend of number no. of tv shows release by year


```
#  trend of number no. of movies release by year
release_year = df_movies.groupby("release_year").agg({"show_id": "nunique"})[df_movies.groupby("release_year").agg({"show_id": "nunique"}).index > 2010]
sns.lineplot(data=release_year, x='release_year', y='show_id', color = "c")
plt.ylabel("Movies Released in the Year")
plt.xlabel("Year")
plt.title("Trend of number no. of Movies release by year")
plt.show()
```

Trend of number no. of Movies release by year


```
#  trend of number no. of movies added to netflix by year
added_year = df_movies.groupby("date_added_year").agg({"show_id": "nunique"})[df_movies.groupby("date_added_year").agg({"show_id": "nunique"}).index > 2010]
sns.lineplot(data=added_year, x='date_added_year', y='show_id', color = "r")
plt.ylabel("Movies added to Netflix by Years")
plt.xlabel("Year")
plt.title("Trend of number no. of movies added to netflix by year")
plt.show()
```

```
#  trend of number no. of tv shows added to netflix by year
added_year = df_shows.groupby("date_added_year").agg({"show_id": "nunique"})[df_shows.groupby("date_added_year").agg({"show_id": "nunique"}).index > 2010]
sns.lineplot(data=added_year, x='date_added_year', y='show_id', color = "r")
plt.ylabel("Movies added to Netflix by Years")
plt.xlabel("Year")
plt.title("Trend of number no. of tv shows added to netflix by year")
plt.show()
```

Trend of number no. of tv shows added to netflix by year

```
df_USA = netflix[netflix["country"] == "United States"]
df_USA_1 = df_USA.groupby("listed_in",as_index=False).agg({"show_id":"nunique"}).sort_values(by = "show_id", ascending = False)[:20]
df_USA_1
```