

# fair-2-version-2

August 24, 2025

## 0.0.1 1. Importing Libraries

```
[868]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime
import calendar
import warnings as wr
wr.filterwarnings('ignore')
```

## 0.0.2 2. Loading the data

```
[869]: # Read the data in the netflix.csv file
df=pd.read_csv("netflix.csv")
df.head()
```

```
[869]: show_id      type      title      director \
0      s1      Movie      Dick Johnson Is Dead      Kirsten Johnson
1      s2      TV Show      Blood & Water      NaN
2      s3      TV Show      Ganglands      Julien Leclercq
3      s4      TV Show      Jailbirds New Orleans      NaN
4      s5      TV Show      Kota Factory      NaN

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

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

```

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

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

```

### 0.0.3 3. Basic Data Analysis

```
[870]: # shape of the data
df.shape
```

```
[870]: (8807, 12)
```

```
[871]: #To see column names of data
df.columns
```

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

```
[872]: # Get information of datatypes of each column
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
```

```
[873]: #describing the data for only numerical column 'release_year'
df.describe()
```

```
[873]:      release_year
count    8807.000000
mean     2014.180198
std        8.819312
min      1925.000000
25%      2013.000000
50%      2017.000000
75%      2019.000000
max      2021.000000
```

```
[874]: # Change datatype of date_added column from object to datetime datatype
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
```

```
[979]: df['date_added']=pd.to_datetime(df['date_added'],errors='coerce')
df['date_added'].head()
```

```
[979]: 0    2021-09-25
1    2021-09-24
2    2021-09-24
3    2021-09-24
4    2021-09-24
```

Name: date\_added, dtype: datetime64[ns]

```
[980]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 13 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        8807 non-null   object
4   cast            8807 non-null   object
5   country         8807 non-null   object
6   date_added      8709 non-null   datetime64[ns]
7   release_year    8807 non-null   int64
8   rating          8807 non-null   object
9   duration        8807 non-null   object
10  listed_in       8807 non-null   object
11  description     8807 non-null   object
12  added_month     8709 non-null   Int64
dtypes: Int64(1), datetime64[ns](1), int64(1), object(10)
memory usage: 903.2+ KB
```

```
[877]: # Null values in dataset
df.isna().sum()
```

```
[877]: show_id         0
type             0
title            0
director        2634
cast            825
country         831
date_added       98
release_year     0
rating           4
duration         3
listed_in        0
description      0
dtype: int64
```

Insights: we have null values in columns are director:2634, cast:825, country:831, date\_added:10, rating:4, duration:3

```
[878]: df["release_year"].loc[df["release_year"].isna()]
```

```
[878]: Series([], Name: release_year, dtype: int64)
```

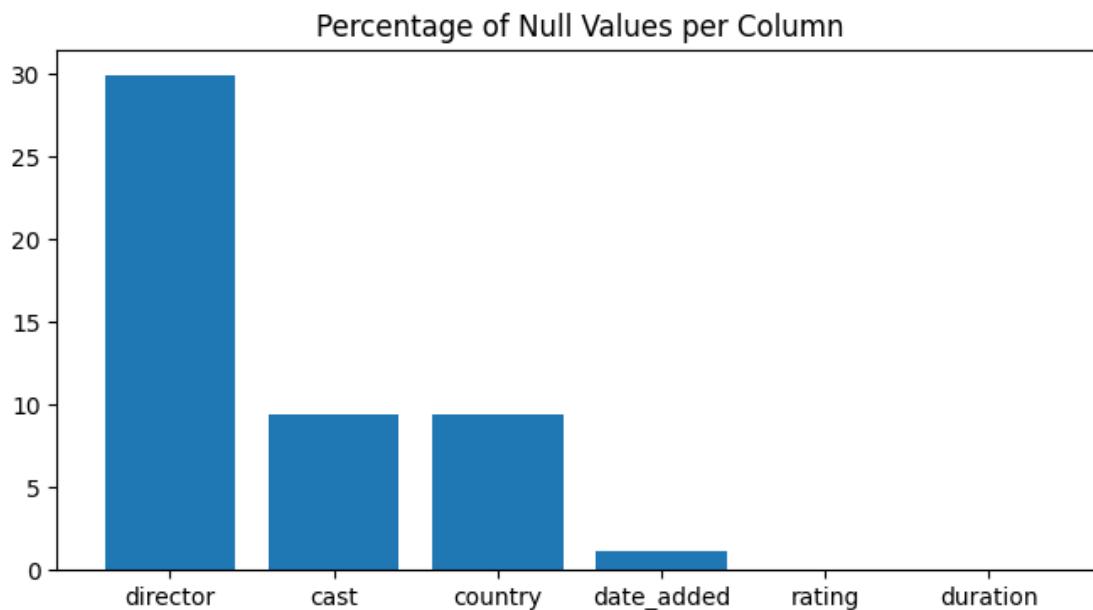
Insights: There are no null values in 'release\_year' column to replace zero.

```
[879]: # Percentage of null values
per_null=(df.isna().sum()*100/len(df))
per_null
per_null_filter = per_null.loc[per_null > 0]
per_null_filter
```

```
[879]: director      29.908028
cast           9.367549
country        9.435676
date_added     1.112751
rating         0.045418
duration       0.034064
dtype: float64
```

Insight: The director column has highest missing values % i.e 30% followed by cast and country as 9% each.

```
[881]: # Plot bar chart
plt.figure(figsize=(8,4))
plt.bar(per_null_filter.index,per_null_filter.values)
plt.title('Percentage of Null Values per Column')
plt.show()
```

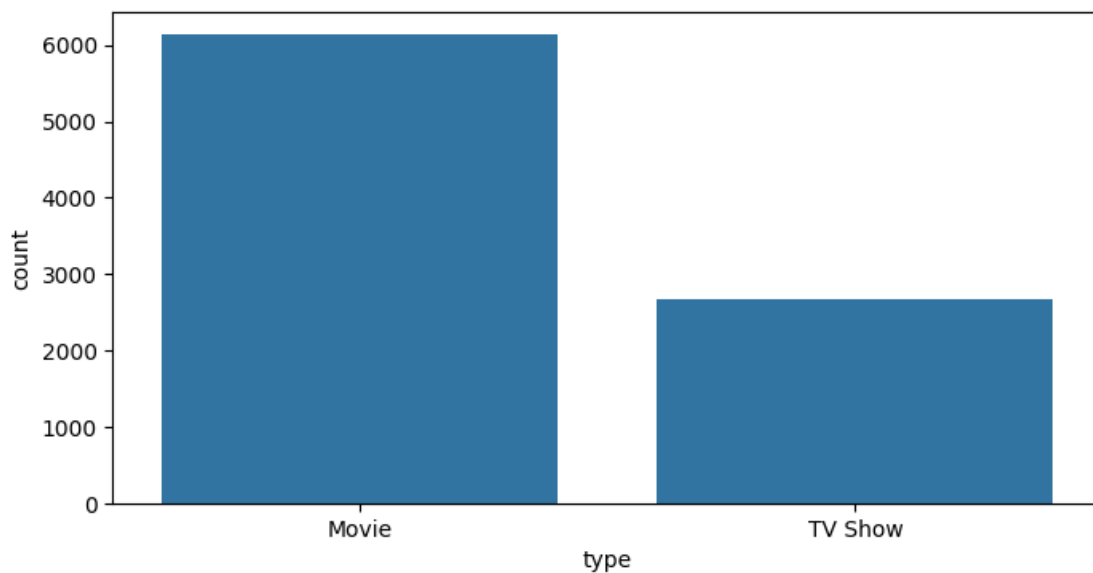


Insight: Percentage of null values are more in directors column followed by cast,country

```
[882]: df["type"].value_counts()
```

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

```
[883]: plt.figure(figsize=(8,4))
sns.countplot(data=df, x='type')
plt.show()
```



Insights: Movies are more than TV Shows in this dataset

```
[884]: #Find null rating values
df.loc[df['rating'].isnull()]
```

```
[884]:
```

	show_id	type	title \
5989	s5990	Movie	13TH: A Conversation with Oprah Winfrey & Ava ...
6827	s6828	TV Show	Gargantia on the Verdurous Planet
7312	s7313	TV Show	Little Lunch
7537	s7538	Movie	My Honor Was Loyalty

	director	cast \
5989	NaN	Oprah Winfrey, Ava DuVernay
6827	NaN	Kaito Ishikawa, Hisako Kanemoto, Ai Kayano, Ka...
7312	NaN	Flynn Curry, Olivia Deeble, Madison Lu, Oisín ...
7537	Alessandro Pepe	Leone Frisa, Paolo Vaccarino, Francesco Miglio...

	country	date_added	release_year	rating	duration	\
5989	NaN	2017-01-26	2017	NaN	37 min	
6827	Japan	2016-12-01	2013	NaN	1 Season	
7312	Australia	2018-02-01	2015	NaN	1 Season	
7537	Italy	2017-03-01	2015	NaN	115 min	

	listed_in	\
5989	Movies	
6827	Anime Series, International TV Shows	
7312	Kids' TV, TV Comedies	
7537	Dramas	

	description
5989	Oprah Winfrey sits down with director Ava DuVe...
6827	After falling through a wormhole, a space-dwel...
7312	Adopting a child's perspective, this show take...
7537	Amid the chaos and horror of World War II, a c...

Insight: There are only 4 rows which has null values in rating column

```
[885]: df['date_added'].isna().sum()
```

```
[885]: np.int64(98)
```

```
[886]: # Unique values in dataset
df.nunique()
```

```
[886]: show_id      8807
type              2
title            8807
director        4528
cast            7692
country          748
date_added      1699
release_year      74
rating           17
duration         220
listed_in        514
description      8775
dtype: int64
```

Insight: Unique values are same for show\_id and title

```
[887]: # Checking unique values for director column
df['director'].nunique()
```

```
[887]: 4528
```

Insights: We have 4528 unique values in director column

```
[888]: # Checking unique values for type column  
df['type'].nunique()
```

```
[888]: 2
```

Insight: There is only one numerical column 'release\_year' in this dataset

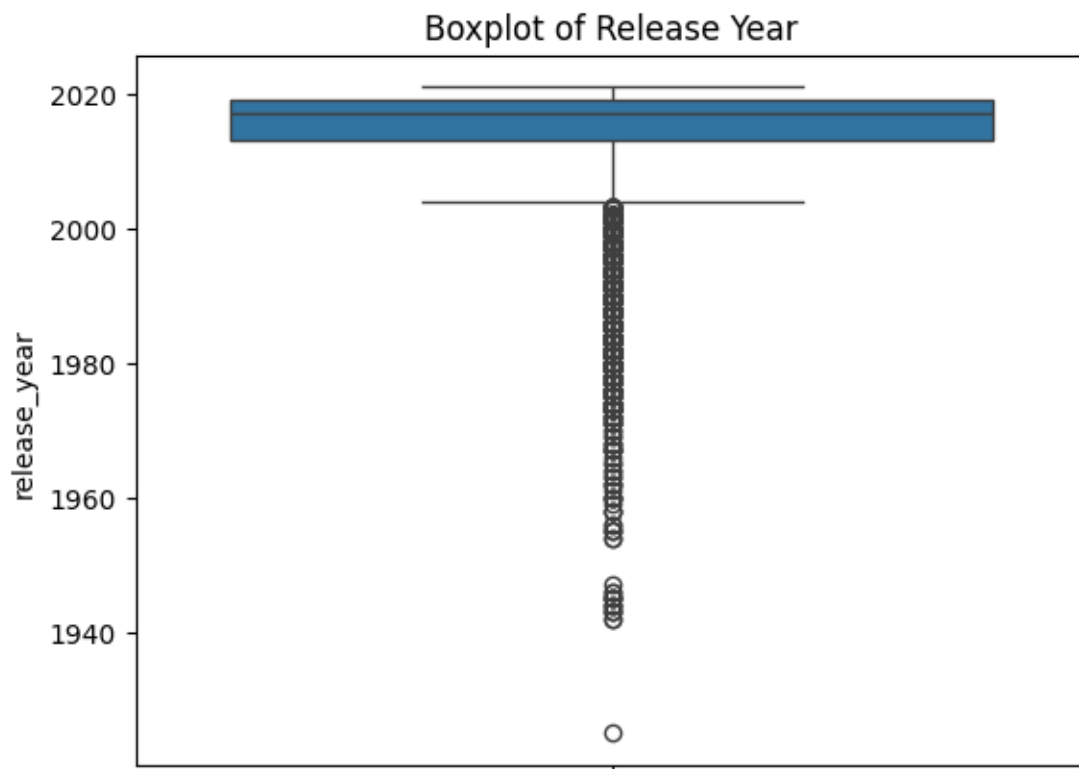
```
[889]: # Is it only Movies and TV-Shows or something else  
df.groupby('type').ngroups
```

```
[889]: 2
```

Insight: There are only Movie and TV Show Groups

#### 0.0.4 Showing Outliers

```
[890]: sns.boxplot(df['release_year'])  
plt.title('Boxplot of Release Year')  
plt.show()
```





Insights: Outliers are more on negative/lower side i.e, below the minimum value-> between 1925 to approx. around 2000. There are no outliers on top of the maximum value.

```
[981]: df.head()
```

```
[981]:  show_id    type          title      director \
0      s1      Movie  Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show      Blood & Water  Unknown Director
2      s3  TV Show      Ganglands    Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans  Unknown Director
4      s5  TV Show      Kota Factory  Unknown Director

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

  date_added  release_year  rating  duration \
0  2021-09-25          2020  PG-13    90 min
1  2021-09-24          2021  TV-MA  2 Seasons
2  2021-09-24          2021  TV-MA  1 Season
3  2021-09-24          2021  TV-MA  1 Season
4  2021-09-24          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  added_month
0  As her father nears the end of his life, filmm...      9
1  After crossing paths at a party, a Cape Town t...      9
2  To protect his family from a powerful drug lor...      9
3  Feuds, flirtations and toilet talk go down amo...      9
4  In a city of coaching centers known to train I...      9
```

categorical columns: show\_id, type, title, director, cast, country, date\_added, rating, duration, listed\_in, description

Numerical Columns: release\_year

```
[892]: #categorical columns:
        ↪ show_id, type, title, director, cast, country, date_added, rating, duration, listed_in, description
        #Numerical Columns: release_year
```

#### 0.0.5 4. Handling null values

For categorical variables with null values, update those rows as unknown column\_name. Replace with 0 for continuous variables having null values.

```
[893]: # Null values in dataset
df.isna().sum()
```

```
[893]: show_id      0
      type        0
      title       0
      director    2634
      cast        825
      country     831
      date_added   98
      release_year 0
      rating       4
      duration     3
      listed_in    0
      description  0
      dtype: int64
```

```
[974]: # Categorical values
cat_columns=["type","title","director","cast","country","rating","duration","listed_in","description"]
for item in cat_columns:
    if item in df.columns:
        df[item]=df[item].fillna(f"Unknown {item.capitalize()}")
```

```
[975]: #Numerical values
df["release_year"] = pd.to_numeric(df["release_year"]).fillna(0).astype(int)
```

```
[976]: df["release_year"].loc[df["release_year"].isna()]
```

```
[976]: Series([], Name: release_year, dtype: int64)
```

```
[982]: df.head()
```

```
[982]: show_id  type      title      director \
0      s1   Movie  Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show      Blood & Water  Unknown Director
2      s3  TV Show      Ganglands     Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans  Unknown Director
4      s5  TV Show      Kota Factory  Unknown Director

      cast      country \
0      Unknown Cast  United States
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...  South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...  Unknown Country
```

3		Unknown Cast	Unknown Country
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...		India

	date_added	release_year	rating	duration	\
0	2021-09-25	2020	PG-13	90 min	
1	2021-09-24	2021	TV-MA	2 Seasons	
2	2021-09-24	2021	TV-MA	1 Season	
3	2021-09-24	2021	TV-MA	1 Season	
4	2021-09-24	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	added_month
0	As her father nears the end of his life, filmm...	9
1	After crossing paths at a party, a Cape Town t...	9
2	To protect his family from a powerful drug lor...	9
3	Feuds, flirtations and toilet talk go down amo...	9
4	In a city of coaching centers known to train I...	9

```
[898]: df['listed_in'].values
```

```
[898]: array(['Documentaries', 'International TV Shows, TV Dramas, TV Mysteries',
        'Crime TV Shows, International TV Shows, TV Action & Adventure',
        ..., 'Comedies, Horror Movies',
        'Children & Family Movies, Comedies',
        'Dramas, International Movies, Music & Musicals'], dtype=object)
```

```
[899]: df['release_year']=pd.to_numeric(df["release_year"]).fillna(0).astype(int)
df['release_year']
```

```
[899]: 0      2020
      1      2021
      2      2021
      3      2021
      4      2021
      ...
     8802    2007
     8803    2018
     8804    2009
     8805    2006
     8806    2015
      Name: release_year, Length: 8807, dtype: int64
```

```
[900]: # Null values in dataset
df.isna().sum()
```

```
[900]: show_id      0
type          0
title         0
director      0
cast          0
country       0
date_added    0
release_year  0
rating        0
duration      0
listed_in     0
description   0
dtype: int64
```

Insight: Data is cleaned. There are no null values in the dataset

#### 0.0.6 4. Un-nesting the columns

Un-nest the columns those have cells with multiple comma separated values by creating multiple rows

```
[983]: df1=df
df1.head()
```

```
[983]: show_id      type      title      director \
0      s1      Movie      Dick Johnson Is Dead      Kirsten Johnson
1      s2      TV Show      Blood & Water      Unknown Director
2      s3      TV Show      Ganglands      Julien Leclercq
3      s4      TV Show      Jailbirds New Orleans      Unknown Director
4      s5      TV Show      Kota Factory      Unknown Director

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

date_added  release_year  rating  duration \
0  2021-09-25      2020  PG-13      90 min
1  2021-09-24      2021  TV-MA  2 Seasons
2  2021-09-24      2021  TV-MA  1 Season
3  2021-09-24      2021  TV-MA  1 Season
4  2021-09-24      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  added_month
0  As her father nears the end of his life, filmm...      9
1  After crossing paths at a party, a Cape Town t...      9
2  To protect his family from a powerful drug lor...      9
3  Feuds, flirtations and toilet talk go down amo...      9
4  In a city of coaching centers known to train I...      9

```

```

[987]: df_exploded = df1.assign(
        director = df1['director'].fillna('Unknown Director').str.split(','),
        cast = df1['cast'].fillna('Unknown Cast').str.split(','),
        country = df1['country'].fillna('Unknown Country').str.split(','),
        listed_in = df1['listed_in'].str.split(',').explode("director").
        explode("cast").explode("country").explode("listed_in")

        print(df_exploded.shape)

```

(202065, 13)

Insights: The multiple comma separated values will creating multiple rows such that each value is present in each row. The rows are increased to 202065. But actual rows in original data is 8807

```

[988]: df_exploded_reset=df_exploded.reset_index()
        df_exploded_reset.head()

```

```

[988]:   index  show_id   type      title      director \
0      0      s1  Movie  Dick Johnson Is Dead  Kirsten Johnson
1      1      s2  TV Show      Blood & Water  Unknown Director
2      1      s2  TV Show      Blood & Water  Unknown Director
3      1      s2  TV Show      Blood & Water  Unknown Director
4      1      s2  TV Show      Blood & Water  Unknown Director

        cast      country  date_added  release_year  rating  duration \
0  Unknown Cast  United States  2021-09-25      2020  PG-13      90 min
1    Ama Qamata  South Africa  2021-09-24      2021  TV-MA  2 Seasons
2    Ama Qamata  South Africa  2021-09-24      2021  TV-MA  2 Seasons
3    Ama Qamata  South Africa  2021-09-24      2021  TV-MA  2 Seasons
4   Khosi Ngema  South Africa  2021-09-24      2021  TV-MA  2 Seasons

        listed_in      description \
0      Documentaries  As her father nears the end of his life, filmm...
1  International TV Shows  After crossing paths at a party, a Cape Town t...

```

2	TV Dramas	After crossing paths at a party, a Cape Town t...
3	TV Mysteries	After crossing paths at a party, a Cape Town t...
4	International TV Shows	After crossing paths at a party, a Cape Town t...

	added_month
0	9
1	9
2	9
3	9
4	9

```
[996]: # removing index column
df_exploded_final=df_exploded_reset.drop("index",axis=1)
df_exploded_final.shape
```

[996]: (202065, 13)

Insights: No. of rows are increased in df\_exploded\_final dataset.

## 1 Find the counts of each categorical and numerical/continuous columns

a. For Non-graphical Analysis:

```
[905]: df["type"].value_counts()
```

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

```
[906]: df["director"].unique()
```

```
[906]: array(['Kirsten Johnson', 'Unknown Director', 'Julien Leclercq', ...,
          'Majid Al Ansari', 'Peter Hewitt', 'Mozes Singh'], dtype=object)
```

```
[907]: df["director"].nunique()
```

[907]: 4529

```
[908]: df["cast"].nunique()
```

[908]: 7693

```
[909]: df["country"].nunique()
```

[909]: 749

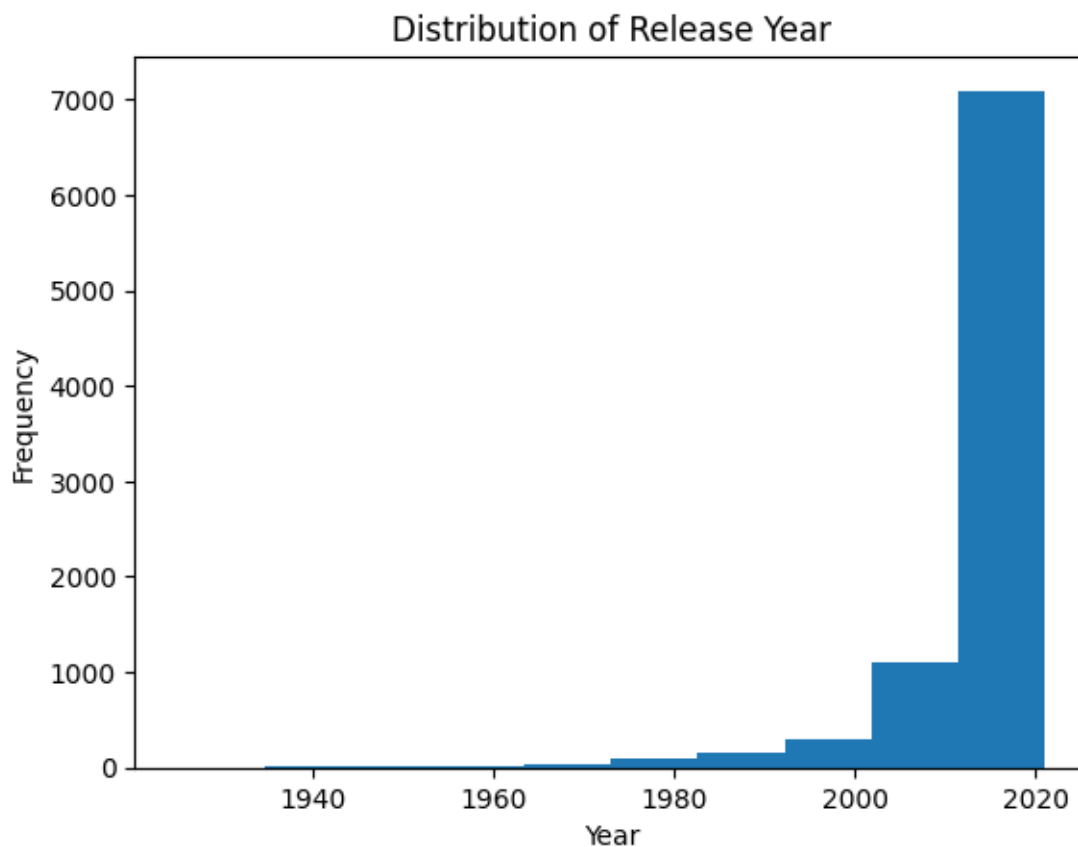
```
[910]: df["listed_in"].nunique()
```

```
[910]: 514
```

b. For graphical analysis:

### 1.0.1 Univariate

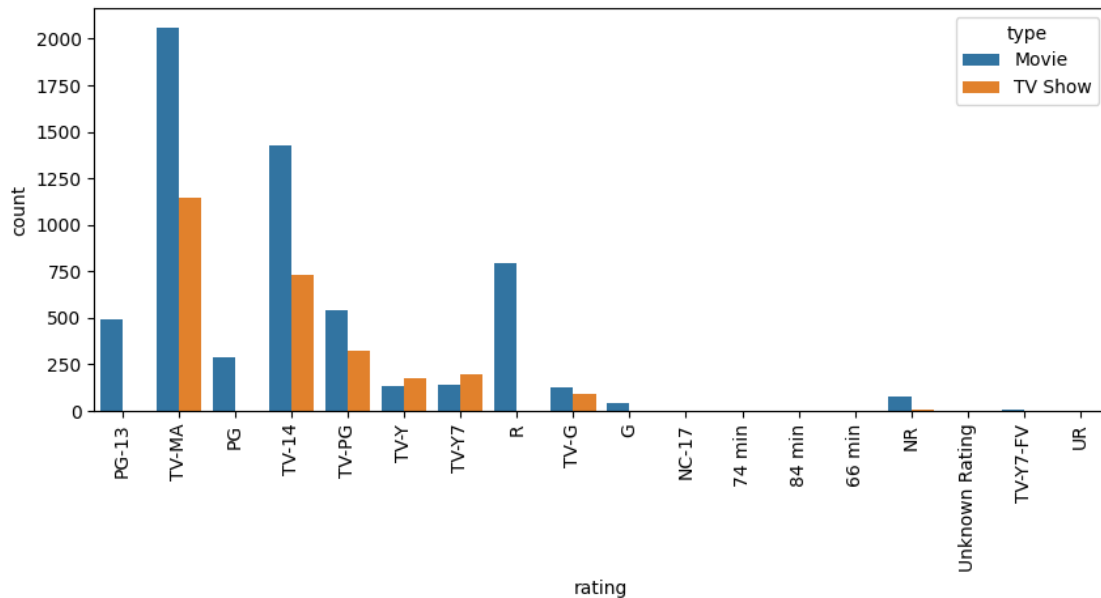
```
[911]: # To know the distribution of release_year column
plt.hist(df["release_year"])
plt.title("Distribution of Release Year")
plt.xlabel("Year")
plt.ylabel("Frequency")
plt.show()
```



Insights: There is spike around 2000,2020 years. It indicates there is lot of content is added to Netflix. It may effect the Netflix revenue.

### 1.0.2 Bivariate

```
[912]: plt.figure(figsize=[10,4])
sns.countplot(x='rating',hue='type',data=df)
plt.xticks(rotation=90)
plt.show()
```

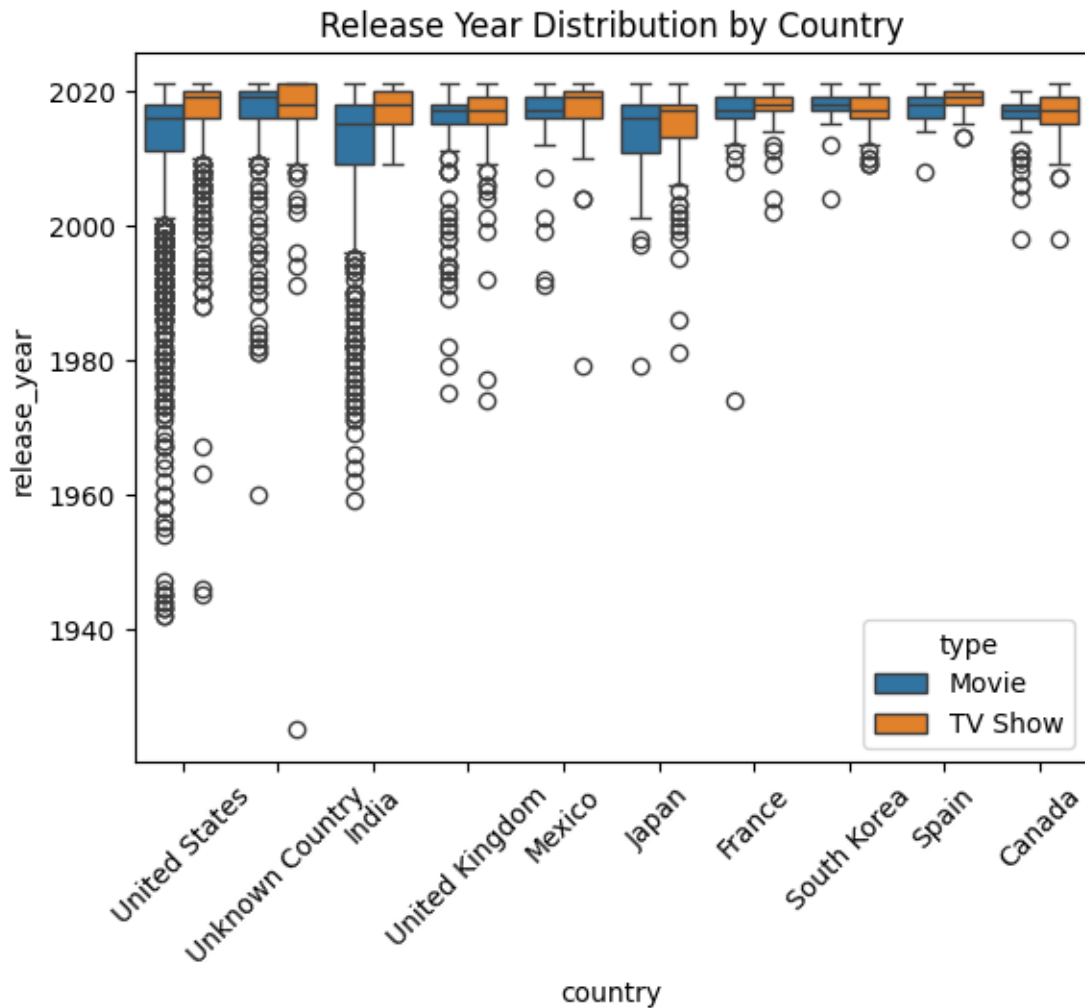


Insight: Customers are watching more content having rating 'TV-MA'

### 1.0.3 Multivariate

```
[913]: # Distribution of top 10 countries in which years and of which type either
        ↪ moves or Tv shows
top_countries = df['country'].value_counts().nlargest(10).index
filtered_df = df[df['country'].isin(top_countries)]
sns.boxplot(x='country', y='release_year', hue='type', data=filtered_df)
plt.xticks(rotation=45)
plt.title("Release Year Distribution by Country")
plt.show()
```





Insights: India Tv Shows have no outliers. India's median is around 2019. So most of the movies are recent.

Egypt has lower median. those movies are around 2010

```
[989]: # Divide the dataset into 2 parts movies, TV Show
movies_df = df.loc[df['type'] == 'Movie']
tvshows_df = df.loc[df['type'] == 'TV Show']
```

```
[990]: movies_df.head()
```

```
[990]:   show_id  type      title \
0      s1  Movie  Dick Johnson Is Dead
6      s7  Movie  My Little Pony: A New Generation
7      s8  Movie      Sankofa
9     s10  Movie    The Starling
```

```

12      s13  Movie                      Je Suis Karl

                                director \
0                                Kirsten Johnson
6      Robert Cullen, José Luis Ucha
7                                Haile Gerima
9                                Theodore Melfi
12     Christian Schwochow

                                cast \
0                                Unknown Cast
6      Vanessa Hudgens, Kimiko Glenn, James Marsden, ...
7      Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
9      Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...
12     Luna Wedler, Jannis Niewöhner, Milan Peschel, ...

                                country date_added \
0                                United States 2021-09-25
6                                Unknown Country 2021-09-24
7      United States, Ghana, Burkina Faso, United Kin... 2021-09-24
9                                United States 2021-09-24
12     Germany, Czech Republic 2021-09-23

release_year rating duration \
0      2020   PG-13    90 min
6      2021    PG     91 min
7      1993   TV-MA   125 min
9      2021   PG-13   104 min
12     2021   TV-MA   127 min

                                listed_in \
0                                Documentaries
6                                Children & Family Movies
7      Dramas, Independent Movies, International Movies
9                                Comedies, Dramas
12     Dramas, International Movies

                                description  added_month
0      As her father nears the end of his life, filmm...      9
6      Equestria's divided. But a bright-eyed hero be...      9
7      On a photo shoot in Ghana, an American model s...      9
9      A woman adjusting to life after a loss contend...      9
12     After most of her family is murdered in a terr...      9

```

```
[991]: tvshows_df.head()
```

```
[991]: show_id      type              title      director \
1      s2  TV Show      Blood & Water  Unknown Director
2      s3  TV Show      Ganglands    Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans  Unknown Director
4      s5  TV Show      Kota Factory  Unknown Director
5      s6  TV Show      Midnight Mass    Mike Flanagan

                                cast      country \
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...  South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...  Unknown Country
3                                Unknown Cast  Unknown Country
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...  India
5  Kate Siegel, Zach Gilford, Hamish Linklater, H...  Unknown Country

date_added  release_year  rating  duration \
1 2021-09-24      2021  TV-MA  2 Seasons
2 2021-09-24      2021  TV-MA  1 Season
3 2021-09-24      2021  TV-MA  1 Season
4 2021-09-24      2021  TV-MA  2 Seasons
5 2021-09-24      2021  TV-MA  1 Season

                                listed_in \
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

description  added_month
1  After crossing paths at a party, a Cape Town t...      9
2  To protect his family from a powerful drug lor...      9
3  Feuds, flirtations and toilet talk go down amo...      9
4  In a city of coaching centers known to train I...      9
5  The arrival of a charismatic young priest brin...      9
```

```
[992]: # Finding oldest movie
movies_df = df.loc[df['type'] == 'Movie']
oldest_movie = movies_df[movies_df['release_year'] == movies_df['release_year'].
    ↪min()]
oldest_movie.copy()
```

```
[992]: show_id      type              title      director \
7790  s7791  Movie      Prelude to War  Frank Capra
8205  s8206  Movie  The Battle of Midway    John Ford

                                cast      country  date_added  release_year \
7790                                Unknown Cast  United States  2017-03-31      1942
```

8205	Henry Fonda, Jane Darwell	United States	2017-03-31	1942
------	---------------------------	---------------	------------	------

	rating	duration	listed_in \
7790	TV-14	52 min	Classic Movies, Documentaries
8205	TV-14	18 min	Classic Movies, Documentaries

		description	added_month
7790	Frank Capra's documentary chronicles the rise ...		3
8205	Director John Ford captures combat footage of ...		3

Insights: These movies(Prelude to War,The Battle of Midway) are the two old movies released in year 1942

```
[993]: # Finding oldest TV Show
tvshows_df = df.loc[df['type'] == 'TV Show']
oldest_tvshow = tvshows_df[tvshows_df['release_year'] ==
    ↳tvshows_df['release_year'].min()]
oldest_tvshow
```

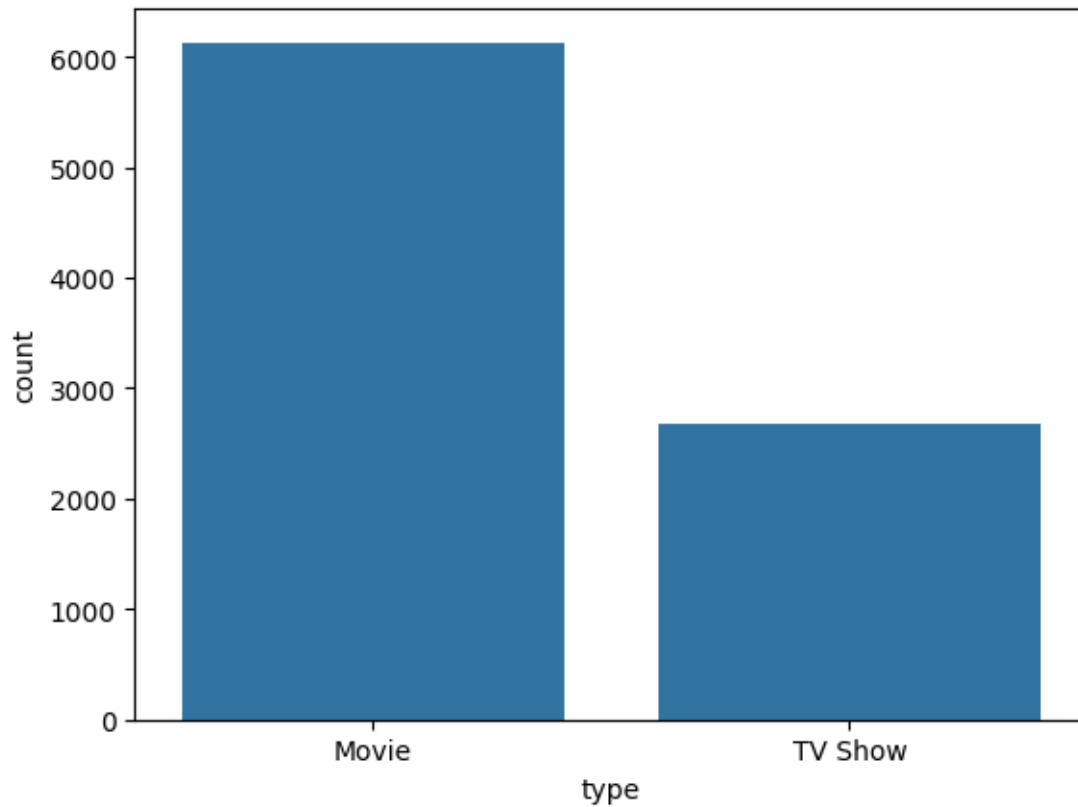
```
[993]:      show_id      type      title      director \
4250    s4251  TV Show  Pioneers: First Women Filmmakers*  Unknown Director

      cast      country date_added  release_year rating  duration \
4250  Unknown Cast  Unknown Country 2018-12-30      1925  TV-14    1 Season

      listed_in      description  added_month
4250  TV Shows  This collection restores films from women who ...      12
```

Insight: This(Pioneers: First Women Filmmakers) TV Show is too old and is released in year 1925

```
[919]: sns.countplot(data=df, x='type')
plt.show()
```



Insights: Customers see more Movies than TV Shows

```
[920]: df['title'].value_counts()
```

```
[920]: title
Zubaan                                1
Dick Johnson Is Dead                  1
Blood & Water                         1
Ganglands                            1
Jailbirds New Orleans                 1
..
Bangkok Breaking                      1
Vendetta: Truth, Lies and The Mafia  1
The Starling                         1
The Great British Baking Show        1
Sankofa                              1
Name: count, Length: 8807, dtype: int64
```

Insights: There are unique titles in title column

```
[921]: df['country'].value_counts()
```

```
[921]: country
United States      2818
India              972
Unknown Country    831
United Kingdom     419
Japan              245
...
Germany, Canada, United States    1
Denmark, United Kingdom, South Africa, Sweden, Belgium    1
Serbia, South Korea, Slovenia    1
United Kingdom, Spain, Belgium    1
South Africa, United States, Japan    1
Name: count, Length: 749, dtype: int64
```

Insights: More movies & TV Shows are from United States followed by India.

```
[922]: #Top 5 movies & TV Shows are from which country
df['country'].value_counts().head()
```

```
[922]: country
United States      2818
India              972
Unknown Country    831
United Kingdom     419
Japan              245
Name: count, dtype: int64
```

Insights: More movies & TV Shows are from United States followed by India,United Kingdom,Japan,South Korea.

```
[923]: #### rating
df['rating'].value_counts()
```

```
[923]: 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
G           41
TV-Y7-FV     6
Unknown Rating  4
NC-17        3
```

```

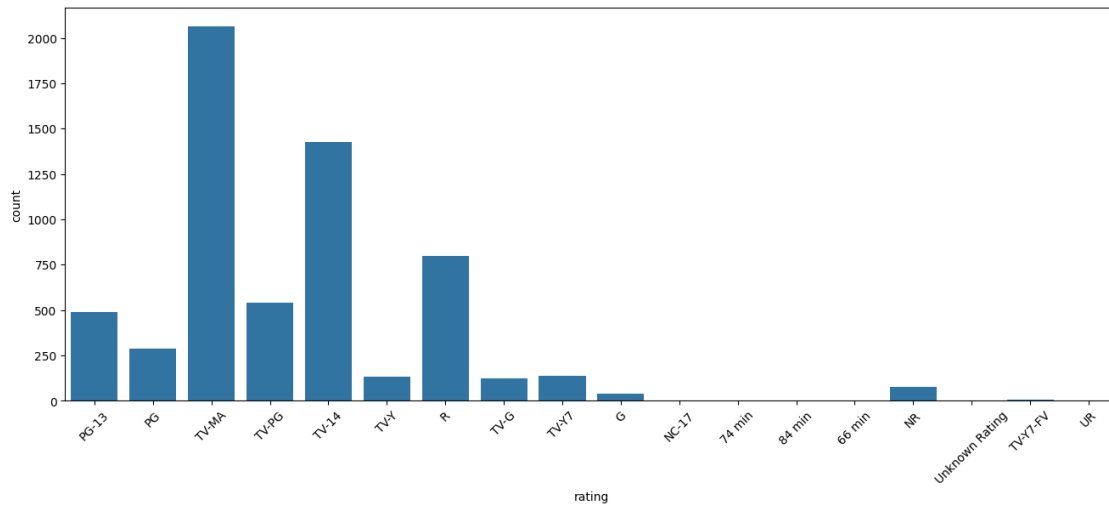
UR          3
66 min      1
74 min      1
84 min      1
Name: count, dtype: int64

```

```

[924]: plt.figure(figsize=(16,6))
sns.countplot(data=movies_df, x='rating')
plt.xticks(rotation=45)
plt.show()

```



Insights: Netflix shows maximum the matured content whose rating is TV-MA than thrillers, dramas and comedies.

```

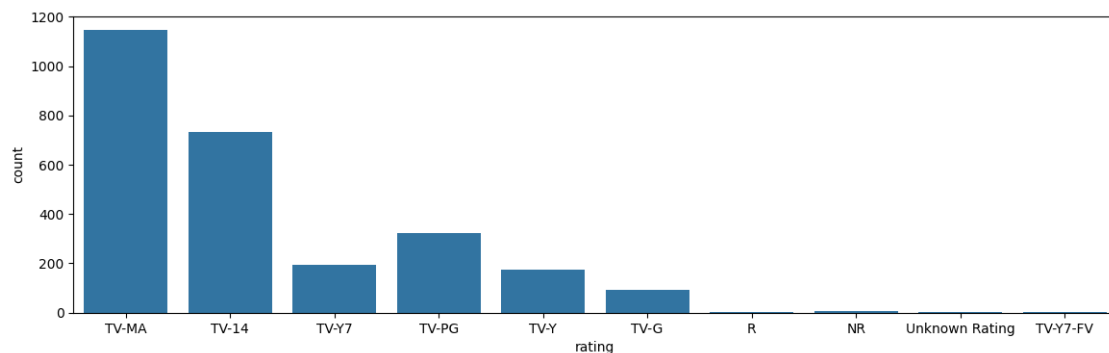
[925]: plt.figure(figsize=(14,4))
sns.countplot(data=tvshows_df, x='rating')

```

```

[925]: <Axes: xlabel='rating', ylabel='count'>

```



Insights: Netflix shows maximum the matured content whose rating is TV-MA

```
[994]: df_listed_in=df.assign(listed_in=df['listed_in'].str.split(',')).
      explode("listed_in")
      df_listed_in.head()
```

```
[994]: show_id      type      title      director \
0      s1      Movie  Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show      Blood & Water  Unknown Director
1      s2  TV Show      Blood & Water  Unknown Director
1      s2  TV Show      Blood & Water  Unknown Director
2      s3  TV Show      Ganglands    Julien Leclercq

      cast      country \
0      Unknown Cast  United States
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...  South Africa
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...  South Africa
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...  South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...  Unknown Country

      date_added  release_year  rating  duration      listed_in \
0  2021-09-25      2020  PG-13      90 min      Documentaries
1  2021-09-24      2021  TV-MA  2 Seasons  International TV Shows
1  2021-09-24      2021  TV-MA  2 Seasons      TV Dramas
1  2021-09-24      2021  TV-MA  2 Seasons      TV Mysteries
2  2021-09-24      2021  TV-MA  1 Season      Crime TV Shows

      description  added_month
0  As her father nears the end of his life, filmm...      9
1  After crossing paths at a party, a Cape Town t...      9
1  After crossing paths at a party, a Cape Town t...      9
1  After crossing paths at a party, a Cape Town t...      9
2  To protect his family from a powerful drug lor...      9
```

```
[927]: # count no. of occurrences in listed_in(genre) column
      df_listed_in['listed_in'].value_counts()
```

```
[927]: listed_in
      International Movies      2624
      Dramas      1600
      Comedies      1210
      Action & Adventure      859
      Documentaries      829
      ...
      Romantic Movies      3
      Spanish-Language TV Shows      2
```



```
LGBTQ Movies          1
TV Sci-Fi & Fantasy   1
Sports Movies         1
Name: count, Length: 73, dtype: int64
```

```
[928]: # How many listed_in values in that column
len(df_listed_in['listed_in'].unique())
```

```
[928]: 73
```

```
[929]: # Finding top 5 genre in listed_in column
df_listed_in['listed_in'].value_counts().head()
```

```
[929]: listed_in
International Movies    2624
Dramas                 1600
Comedies               1210
Action & Adventure     859
Documentaries          829
Name: count, dtype: int64
```

Insight: Customers are seeing more International Movies than any other shows

Recommendations: We have to add more International Movies to customers list in Netflix.

```
[997]: # Find movies dataset from the dataset df_exploded_final which is having the
↳ new index and also no multiple comma seperated values
movies_final = df_exploded_final[df_exploded_final['type'] == 'Movie']
movies_final.head()
```

```
[997]:
```

	show_id	type	title	director	\
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	
159	s7	Movie	My Little Pony: A New Generation	Robert Cullen	
160	s7	Movie	My Little Pony: A New Generation	Robert Cullen	
161	s7	Movie	My Little Pony: A New Generation	Robert Cullen	
162	s7	Movie	My Little Pony: A New Generation	Robert Cullen	

	cast	country	date_added	release_year	rating	\
0	Unknown Cast	United States	2021-09-25	2020	PG-13	
159	Vanessa Hudgens	Unknown Country	2021-09-24	2021	PG	
160	Kimiko Glenn	Unknown Country	2021-09-24	2021	PG	
161	James Marsden	Unknown Country	2021-09-24	2021	PG	
162	Sofia Carson	Unknown Country	2021-09-24	2021	PG	

	duration	listed_in	\
0	90 min	Documentaries	
159	91 min	Children & Family Movies	
160	91 min	Children & Family Movies	

```
161 91 min Children & Family Movies
162 91 min Children & Family Movies
```

	description	added_month
0	As her father nears the end of his life, filmm...	9
159	Equestria's divided. But a bright-eyed hero be...	9
160	Equestria's divided. But a bright-eyed hero be...	9
161	Equestria's divided. But a bright-eyed hero be...	9
162	Equestria's divided. But a bright-eyed hero be...	9

```
[931]: movies_final['director'].nunique()
```

```
[931]: 4887
```

```
[932]: movies_final.groupby('director').ngroups
```

```
[932]: 4887
```

```
[933]: # Get the rows of director named Aaron Moorhead
movies_final.groupby('director').get_group(' Aaron Moorhead')

# Find no. of unique directors who made movies
movies_final.groupby('director')['title'].nunique().nlargest(10)
```

```
[933]: director
Unknown Director      188
Rajiv Chilaka         22
Jan Suter             18
Raúl Campos           18
Suhas Kadav           16
Jay Karas             15
Marcus Raboy          15
Cathy Garcia-Molina   13
Jay Chapman           12
Martin Scorsese       12
Name: title, dtype: int64
```

Insight: There is no name(Unknown Director) for directors who made maximum no. of movies followed by 'Rajiv Chilaka'

```
[934]: # Find all the counties who are watching movies
movies_final['country'].value_counts()
```

```
[934]: country
United States      38316
India              20761
United States       7501
```

Unknown Country	6199
United Kingdom	5655
...	
Botswana	2
Afghanistan	2
Nicaragua	1
Kazakhstan	1
Uganda	1

Name: count, Length: 188, dtype: int64

```
[935]: # Find the country who is watching more movies
movies_final['country'].value_counts().max()
```

[935]: 38316

Insights: Highest movies are watched by customers in the United States country

```
[936]: # Find the movies names and its count done by each director
movies_final.groupby('director')['title'].value_counts()
```

```
[936]: director      title
Aaron Moorhead    The Endless      18
                  Synchronic        6
Aaron Woolf       To Be a Miss      2
Abbas Alibhai Burmawalla Race      18
Abdullah Al Noor  Sincerely Yours, Dhaka 30
                  ..
Óskar Thór Axelsson  Ég man þig      20
Ömer Faruk Sorak    G.O.R.A         30
                  Aşk Tesadüfleri Sever 24
Şenol Sönmez       Hayat Öpücüğü    24
                  Kill Me If You Dare  24
Name: count, Length: 6854, dtype: int64
```

```
[937]: # Find the no. of movies done by each director
movies_final.groupby('director')['title'].count()
```

```
[937]: director
Aaron Moorhead      24
Aaron Woolf         2
Abbas Alibhai Burmawalla 18
Abdullah Al Noor    30
Abhinav Shiv Tiwari  24
                  ..
Çagan Irmak        20
Ísold Uggadóttir    27
Óskar Thór Axelsson 20
```

```
Ömer Faruk Sorak          54
Şenol Sönmez              48
Name: title, Length: 4887, dtype: int64
```

## 2 Questions

1. How has the number of movies released per year changed over the last 20-30 years?

```
[938]: movies_df['release_year'].max()
```

```
[938]: 2021
```

```
[939]: # Today's date and time
datetime.now()
```

```
[939]: datetime.datetime(2025, 8, 24, 18, 14, 8, 835528)
```

```
[940]: year_30= datetime.now().year - 30
year_30
```

```
[940]: 1995
```

```
[998]: # movies release from year 1995 (max. year is 2025)
recent_movies=movies_df.loc[movies_df['release_year']>= year_30]
recent_movies.head()
```

```
[998]:
```

	show_id	type	title \
0	s1	Movie	Dick Johnson Is Dead
6	s7	Movie	My Little Pony: A New Generation
9	s10	Movie	The Starling
12	s13	Movie	Je Suis Karl
13	s14	Movie	Confessions of an Invisible Girl

	director \
0	Kirsten Johnson
6	Robert Cullen, José Luis Ucha
9	Theodore Melfi
12	Christian Schwochow
13	Bruno Garotti

	cast \
0	Unknown Cast
6	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...
9	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...
12	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...
13	Klara Castanho, Lucca Picon, Júlia Gomes, Marc...

	country	date_added	release_year	rating	duration	\
0	United States	2021-09-25	2020	PG-13	90 min	
6	Unknown Country	2021-09-24	2021	PG	91 min	
9	United States	2021-09-24	2021	PG-13	104 min	
12	Germany, Czech Republic	2021-09-23	2021	TV-MA	127 min	
13	Unknown Country	2021-09-22	2021	TV-PG	91 min	

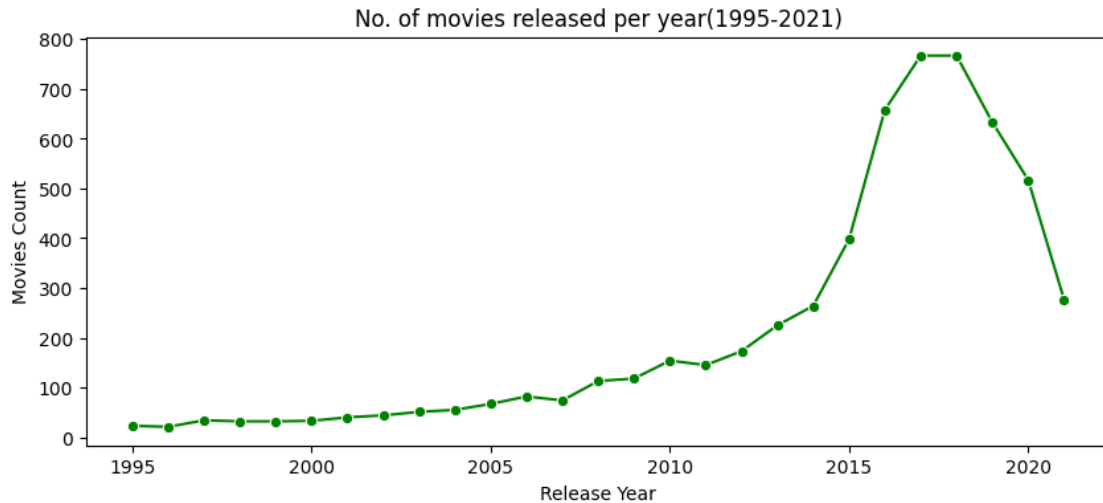
	listed_in	\
0	Documentaries	
6	Children & Family Movies	
9	Comedies, Dramas	
12	Dramas, International Movies	
13	Children & Family Movies, Comedies	

	description	added_month
0	As her father nears the end of his life, filmm...	9
6	Equestria's divided. But a bright-eyed hero be...	9
9	A woman adjusting to life after a loss contend...	9
12	After most of her family is murdered in a terr...	9
13	When the clever but socially-awkward Tetê join...	9

```
[942]: # count movies per year
movies_per_year=recent_movies['release_year'].value_counts().sort_index()
movies_per_year.head()
```

```
[942]: release_year
1995    23
1996    21
1997    34
1998    32
1999    32
Name: count, dtype: int64
```

```
[943]: plt.figure(figsize=(10,4))
sns.lineplot(x=movies_per_year.index,y=movies_per_year.
             ↪values,marker='o',color='green')
plt.title("No. of movies released per year(1995-2021)")
plt.xlabel('Release Year')
plt.ylabel("Movies Count")
plt.xticks(rotation=0)
plt.show()
```



Insights: Untill 2015 there is steady growth in movies. After 2015 there is high peak noticed ie., more no. of movies are released in 2017,2018 and from there it is decreased till now.

## 2. Comparison of tv shows vs. movies.

```
[999]: movies_df.head()
```

```
[999]:
```

	show_id	type	title \
0	s1	Movie	Dick Johnson Is Dead
6	s7	Movie	My Little Pony: A New Generation
7	s8	Movie	Sankofa
9	s10	Movie	The Starling
12	s13	Movie	Je Suis Karl

	director \
0	Kirsten Johnson
6	Robert Cullen, José Luis Ucha
7	Haile Gerima
9	Theodore Melfi
12	Christian Schwochow

	cast \
0	Unknown Cast
6	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...
7	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
9	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...
12	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...

	country date_added \
0	United States 2021-09-25

6	Unknown Country	2021-09-24
7	United States, Ghana, Burkina Faso, United Kin...	2021-09-24
9	United States	2021-09-24
12	Germany, Czech Republic	2021-09-23

	release_year	rating	duration	\
0	2020	PG-13	90 min	
6	2021	PG	91 min	
7	1993	TV-MA	125 min	
9	2021	PG-13	104 min	
12	2021	TV-MA	127 min	

	listed_in	\
0	Documentaries	
6	Children & Family Movies	
7	Dramas, Independent Movies, International Movies	
9	Comedies, Dramas	
12	Dramas, International Movies	

	description	added_month
0	As her father nears the end of his life, filmm...	9
6	Equestria's divided. But a bright-eyed hero be...	9
7	On a photo shoot in Ghana, an American model s...	9
9	A woman adjusting to life after a loss contend...	9
12	After most of her family is murdered in a terr...	9

```
[945]: year_15= datetime.now().year - 15
year_15
```

```
[945]: 2010
```

```
[946]: # movies release from year 2010 to 2025
recent_movies_15=movies_df.loc[movies_df['release_year']>= year_15]
recent_movies_15.shape
# count movies per year
movies_count=recent_movies_15['release_year'].value_counts().sort_index()
movies_count
```

```
[946]: release_year
2010    154
2011    145
2012    173
2013    225
2014    264
2015    398
2016    658
2017    767
```

```

2018    767
2019    633
2020    517
2021    277
Name: count, dtype: int64

```

```
[1000]: tvshows_df.head()
```

```

[1000]:  show_id      type      title      director \
1      s2  TV Show      Blood & Water  Unknown Director
2      s3  TV Show      Ganglands      Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans  Unknown Director
4      s5  TV Show      Kota Factory  Unknown Director
5      s6  TV Show      Midnight Mass      Mike Flanagan

                                cast      country \
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...  South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...  Unknown Country
3                                Unknown Cast  Unknown Country
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...  India
5  Kate Siegel, Zach Gilford, Hamish Linklater, H...  Unknown Country

  date_added  release_year  rating  duration \
1 2021-09-24      2021  TV-MA  2 Seasons
2 2021-09-24      2021  TV-MA  1 Season
3 2021-09-24      2021  TV-MA  1 Season
4 2021-09-24      2021  TV-MA  2 Seasons
5 2021-09-24      2021  TV-MA  1 Season

                                listed_in \
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

                                description  added_month
1  After crossing paths at a party, a Cape Town t...      9
2  To protect his family from a powerful drug lor...      9
3  Feuds, flirtations and toilet talk go down amo...      9
4  In a city of coaching centers known to train I...      9
5  The arrival of a charismatic young priest brin...      9

```

```

[948]: # tvshows release from year 2010 to 2025
recent_tvshows_15=tvshows_df.loc[tvshows_df['release_year']>= year_15]
tvshows_count=recent_tvshows_15['release_year'].value_counts().sort_index()
tvshows_count

```



```
[948]: release_year
2010    40
2011    40
2012    64
2013    63
2014    88
2015   162
2016   244
2017   265
2018   380
2019   397
2020   436
2021   315
Name: count, dtype: int64
```

```
[949]: comparision_df=pd.DataFrame({
        'Year':movies_count.index,
        'Movies':movies_count.values,
        'TV Shows':tvshows_count.values

    })
comparision_df
comparison_melted = comparision_df.melt(id_vars='Year', value_vars=['Movies',
↪ 'TV Shows'],
                                     var_name='Type', value_name='Count')
comparison_melted
```

```
[949]:
```

	Year	Type	Count
0	2010	Movies	154
1	2011	Movies	145
2	2012	Movies	173
3	2013	Movies	225
4	2014	Movies	264
5	2015	Movies	398
6	2016	Movies	658
7	2017	Movies	767
8	2018	Movies	767
9	2019	Movies	633
10	2020	Movies	517
11	2021	Movies	277
12	2010	TV Shows	40
13	2011	TV Shows	40
14	2012	TV Shows	64
15	2013	TV Shows	63
16	2014	TV Shows	88
17	2015	TV Shows	162
18	2016	TV Shows	244

```

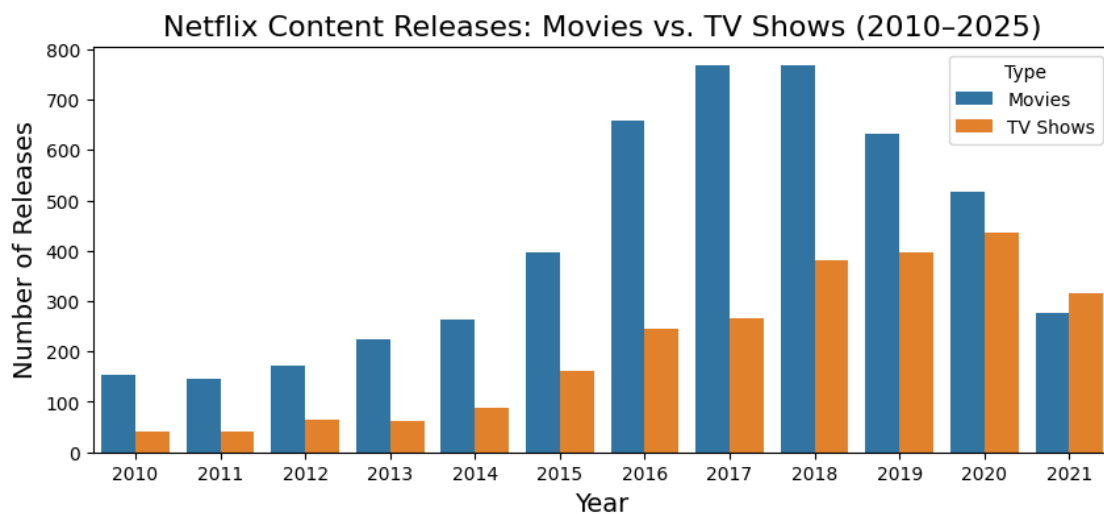
19 2017 TV Shows 265
20 2018 TV Shows 380
21 2019 TV Shows 397
22 2020 TV Shows 436
23 2021 TV Shows 315

```

```

[950]: plt.figure(figsize=(10,4))
sns.barplot(x='Year',y='Count',hue='Type',data=comparison_melted)
plt.title('Netflix Content Releases: Movies vs. TV Shows (2010-2025)',
          fontsize=16)
plt.xlabel('Year', fontsize=14)
plt.ylabel('Number of Releases', fontsize=14)
plt.show()

```



Insights: There is significance increase in TV Shows from 2015 to 2020 and later it is slightly decreased. There is high peak observed from 2014 to 2018 years. Due to pandemic of covid-19 the movies released are gradually decreasing from 2019 to 2021. There is equal distribution in 2017, 2018 years.

### 3. What is the best time to launch a TV show?

```

[951]: tvshows_df.head()

```

```

[951]: show_id    type    title    director \
1      s2  TV Show  Blood & Water  Unknown Director
2      s3  TV Show  Ganglands    Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans  Unknown Director
4      s5  TV Show  Kota Factory  Unknown Director
5      s6  TV Show  Midnight Mass    Mike Flanagan

```

	cast	country \
1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Unknown Country
3	Unknown Cast	Unknown Country
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India
5	Kate Siegel, Zach Gilford, Hamish Linklater, H...	Unknown Country

	date_added	release_year	rating	duration \
1	2021-09-24 00:00:00	2021	TV-MA	2 Seasons
2	2021-09-24 00:00:00	2021	TV-MA	1 Season
3	2021-09-24 00:00:00	2021	TV-MA	1 Season
4	2021-09-24 00:00:00	2021	TV-MA	2 Seasons
5	2021-09-24 00:00:00	2021	TV-MA	1 Season

	listed_in \
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

	description
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...

```
[1001]: tvshows_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2676 entries, 1 to 8803
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
---  -
0   show_id         2676 non-null   object
1   type            2676 non-null   object
2   title           2676 non-null   object
3   director        2676 non-null   object
4   cast            2676 non-null   object
5   country         2676 non-null   object
6   date_added      2578 non-null   datetime64[ns]
7   release_year    2676 non-null   int64
8   rating          2676 non-null   object
9   duration        2676 non-null   object
10  listed_in       2676 non-null   object
11  description     2676 non-null   object
```

```

12 added_month 2578 non-null Int64
dtypes: Int64(1), datetime64[ns](1), int64(1), object(10)
memory usage: 295.3+ KB

```

```

[1002]: df_months=df
df_months.head()

```

```

[1002]: show_id      type      title      director \
0      s1      Movie      Dick Johnson Is Dead      Kirsten Johnson
1      s2      TV Show      Blood & Water      Unknown Director
2      s3      TV Show      Ganglands      Julien Leclercq
3      s4      TV Show      Jailbirds New Orleans      Unknown Director
4      s5      TV Show      Kota Factory      Unknown Director

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

date_added  release_year  rating  duration \
0 2021-09-25      2020  PG-13      90 min
1 2021-09-24      2021  TV-MA  2 Seasons
2 2021-09-24      2021  TV-MA  1 Season
3 2021-09-24      2021  TV-MA  1 Season
4 2021-09-24      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  added_month
0  As her father nears the end of his life, filmm...      9
1  After crossing paths at a party, a Cape Town t...      9
2  To protect his family from a powerful drug lor...      9
3  Feuds, flirtations and toilet talk go down amo...      9
4  In a city of coaching centers known to train I...      9

```

```

[1003]: # Extract month from date
df_months["added_month"] = pd.
↳to_datetime(df_months["date_added"],errors='coerce').dt.month.astype('Int64')
#extract TV Show data from dataset df_months
tv_df=df_months.loc[df_months['type']=='TV Show']

```

```
tv_df.head()
```

```
[1003]:
```

	show_id	type	title	director	\
1	s2	TV Show	Blood & Water	Unknown Director	
2	s3	TV Show	Ganglands	Julien Leclercq	
3	s4	TV Show	Jailbirds New Orleans	Unknown Director	
4	s5	TV Show	Kota Factory	Unknown Director	
5	s6	TV Show	Midnight Mass	Mike Flanagan	

	cast	country	\
1	Ama Qamata, Khosi Ngema, Gail Mababane, Thaban...	South Africa	
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Unknown Country	
3	Unknown Cast	Unknown Country	
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	
5	Kate Siegel, Zach Gilford, Hamish Linklater, H...	Unknown Country	

	date_added	release_year	rating	duration	\
1	2021-09-24	2021	TV-MA	2 Seasons	
2	2021-09-24	2021	TV-MA	1 Season	
3	2021-09-24	2021	TV-MA	1 Season	
4	2021-09-24	2021	TV-MA	2 Seasons	
5	2021-09-24	2021	TV-MA	1 Season	

	listed_in	\
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	

	description	added_month
1	After crossing paths at a party, a Cape Town t...	9
2	To protect his family from a powerful drug lor...	9
3	Feuds, flirtations and toilet talk go down amo...	9
4	In a city of coaching centers known to train I...	9
5	The arrival of a charismatic young priest brin...	9

```
[955]: # Find count of tv shows released on Netflix for each month
month_counts_tv=tv_df["added_month"].value_counts().sort_index()
month_counts_tv
```

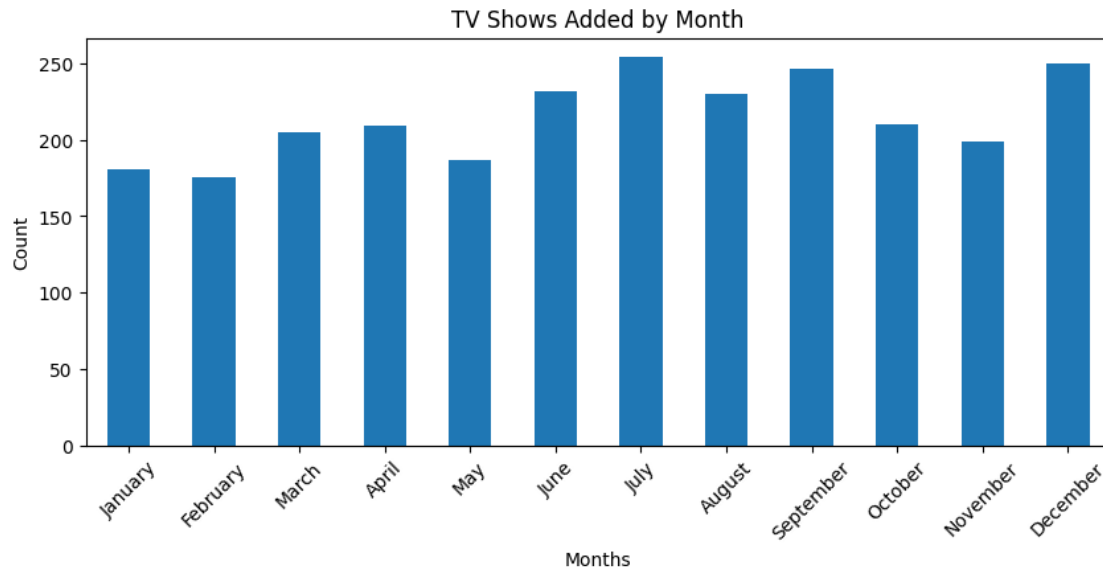
```
[955]: added_month
1      181
2      175
3      205
4      209
5      187
```

```
6      232
7      254
8      230
9      246
10     210
11     199
12     250
Name: count, dtype: Int64
```

```
[956]: month_counts_tv.index = [calendar.month_name[int(m)] for m in month_counts_tv.
    ↪index]
month_counts_tv
```

```
[956]: January      181
February    175
March       205
April       209
May         187
June        232
July        254
August      230
September   246
October     210
November    199
December    250
Name: count, dtype: Int64
```

```
[957]: plt.figure(figsize=(10,4))
month_counts_tv.plot(kind='bar')
plt.title("TV Shows Added by Month")
plt.xlabel("Months")
plt.ylabel("Count")
plt.xticks(rotation=45)
plt.show()
```



Insights: In July month, 234 tv shows are added to Netflix. This is the best time to add to Netflix. Lowest of all months is February.

#### 4. Analysis of actors/directors of different types of shows/movies.

```
[958]: df_actors = df.assign(cast=df["cast"].str.split(",")).explode("cast")
x=df_actors.reset_index()
df_actors=x.drop("index",axis=1)
#df_actors
df_actors["cast"] = df_actors["cast"].str.strip()
#df_actors
df_actors["cast"].value_counts().head(10)
```

```
[958]: cast
Unknown Cast      825
Anupam Kher       43
Shah Rukh Khan    35
Julie Teiwani     33
Naseeruddin Shah  32
Takahiro Sakurai  32
Rupa Bhimani      31
Om Puri           30
Akshay Kumar      30
Yuki Kaji         29
Name: count, dtype: int64
```

```
[959]: df_directors_new = df.assign(director=df["director"].str.split(",")).
        ↪explode("director")
```

```

y=df_directors_new.reset_index()
df_directors_new=y.drop("index",axis=1)
df_directors_new
df_directors_new["director"] = df_directors_new["director"].str.strip()
df_directors_new
df_directors_new["director"].value_counts().head(10)

```

```

[959]: director
Unknown Director      2634
Rajiv Chilaka         22
Jan Suter              21
Raúl Campos           19
Suhas Kadav           16
Marcus Raboy          16
Jay Karas             15
Cathy Garcia-Molina   13
Martin Scorsese       12
Youssef Chahine       12
Name: count, dtype: int64

```

```

[960]: # For each type count actors and directors
top_actors_by_type = (df_actors.groupby("type")["cast"].value_counts().
    ↳groupby(level=0).head(5).reset_index(name="Count"))

top_directors_by_type = (df_directors_new.groupby("type")["director"].
    ↳value_counts().groupby(level=0).head(5).reset_index(name="Count"))

```

```

[961]: top_actors_by_type

```

```

[961]:
   type      cast  Count
0  Movie  Unknown Cast   475
1  Movie   Anupam Kher    42
2  Movie  Shah Rukh Khan   35
3  Movie  Naseeruddin Shah  32
4  Movie   Akshay Kumar   30
5  TV Show  Unknown Cast  350
6  TV Show  Takahiro Sakurai  25
7  TV Show    Yuki Kaji   19
8  TV Show    Ai Kayano   17
9  TV Show   Daisuke Ono   17

```

```

[962]: top_directors_by_type

```

```

[962]:
   type      director  Count
0  Movie  Unknown Director   188
1  Movie    Rajiv Chilaka    22
2  Movie    Jan Suter       21

```



3	Movie	Raúl Campos	19
4	Movie	Suhas Kadav	16
5	TV Show	Unknown Director	2446
6	TV Show	Alastair Fothergill	3
7	TV Show	Ken Burns	3
8	TV Show	Gautham Vasudev Menon	2
9	TV Show	Hsu Fu-chun	2

Insights: Most famous actor's movies are added more in Netflix. Netflix is more concentrated on movies based on popular actors. TV Shows build longer time and actors who are famous those TV shows are added more. Directors who make movies are more when compared to directors of TV Shows.

## 5. Does Netflix has more focus on TV Shows than movies in recent years

```
[963]: df["release_year"] = pd.to_numeric(df["release_year"], errors="coerce").
      ↪ fillna(0).astype(int)
```

```
[1004]: # Focus on the last 10 years
recent_threshold = df["release_year"].max() - 10
df_recent = df.loc[df["release_year"] >= recent_threshold]
df_recent.head()
```

```
[1004]: show_id    type      title      director \
0      s1      Movie  Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show      Blood & Water  Unknown Director
2      s3  TV Show      Ganglands      Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans  Unknown Director
4      s5  TV Show      Kota Factory  Unknown Director

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

date_added  release_year  rating  duration \
0  2021-09-25           2020  PG-13    90 min
1  2021-09-24           2021  TV-MA  2 Seasons
2  2021-09-24           2021  TV-MA  1 Season
3  2021-09-24           2021  TV-MA  1 Season
4  2021-09-24           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	added_month
0	As her father nears the end of his life, filmm...	9
1	After crossing paths at a party, a Cape Town t...	9
2	To protect his family from a powerful drug lor...	9
3	Feuds, flirtations and toilet talk go down amo...	9
4	In a city of coaching centers known to train I...	9

```

[965]: # Count Movies vs TV Shows per year
trend = df_recent.groupby(["release_year", "type"]).size().
        ↪reset_index(name="count")
pivot_trend = trend.pivot(index="release_year", columns="type", values="count").
        ↪fillna(0)
pivot_trend

```

```

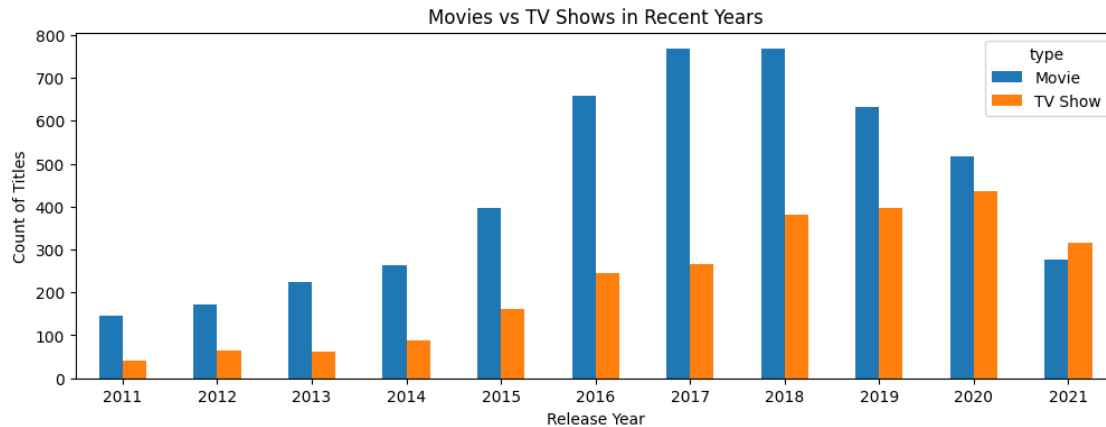
[965]: type      Movie  TV Show
release_year
2011          145      40
2012          173      64
2013          225      63
2014          264      88
2015          398     162
2016          658     244
2017          767     265
2018          767     380
2019          633     397
2020          517     436
2021          277     315

```

```

[966]: # Plot
pivot_trend.plot(kind="bar", figsize=(12,4))
plt.title("Movies vs TV Shows in Recent Years")
plt.xlabel("Release Year")
plt.ylabel("Count of Titles")
plt.xticks(rotation=0)
plt.show()

```



Insights: Netflix is more concentrated on movies from 2015, movies are gradually increasing and goes to peak in 2017,2018. From 2017 to 2020 Netflix has increased it's focus on TV Shows also.

**Which genre movies are more popular or produced more**

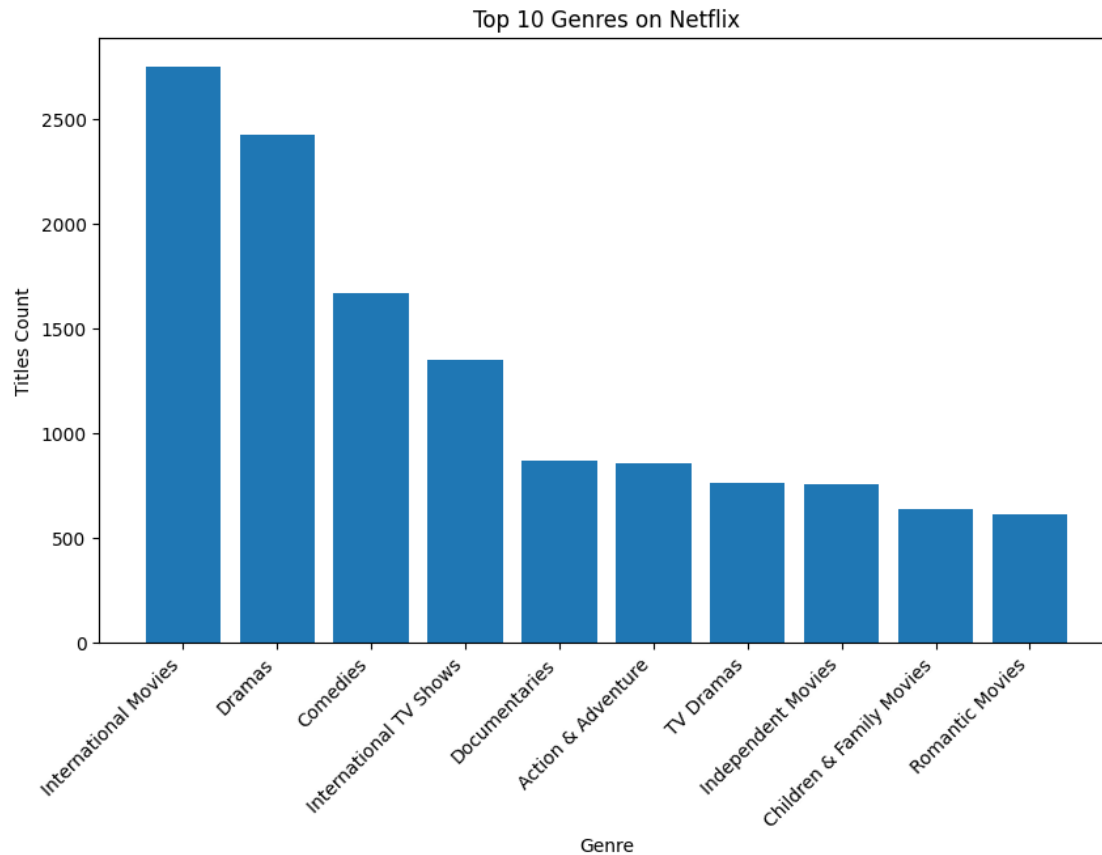
```
[967]: df_genres = df.assign(Genre=df['listed_in'].str.split(', ').explode('Genre'))

# Count how many titles are in each genre
genre_counts = df_genres['Genre'].value_counts().reset_index()
genre_counts.columns = ['Genre', 'Count']
# Display Top 10 genres
genre_counts.head(10)
```

```
[967]:
```

	Genre	Count
0	International Movies	2752
1	Dramas	2427
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

```
[968]: # Plot top 10 genres
plt.figure(figsize=(10,6))
plt.bar(genre_counts['Genre'][:10], genre_counts['Count'][:10])
plt.xticks(rotation=45,ha='right')
plt.title("Top 10 Genres on Netflix")
plt.xlabel("Genre")
plt.ylabel("Titles Count")
plt.show()
```



Insights: The count of International Movies is on higher side when compared to other genres and it is followed by Dramas, comedies.

## 6. Understanding what content is available in different countries

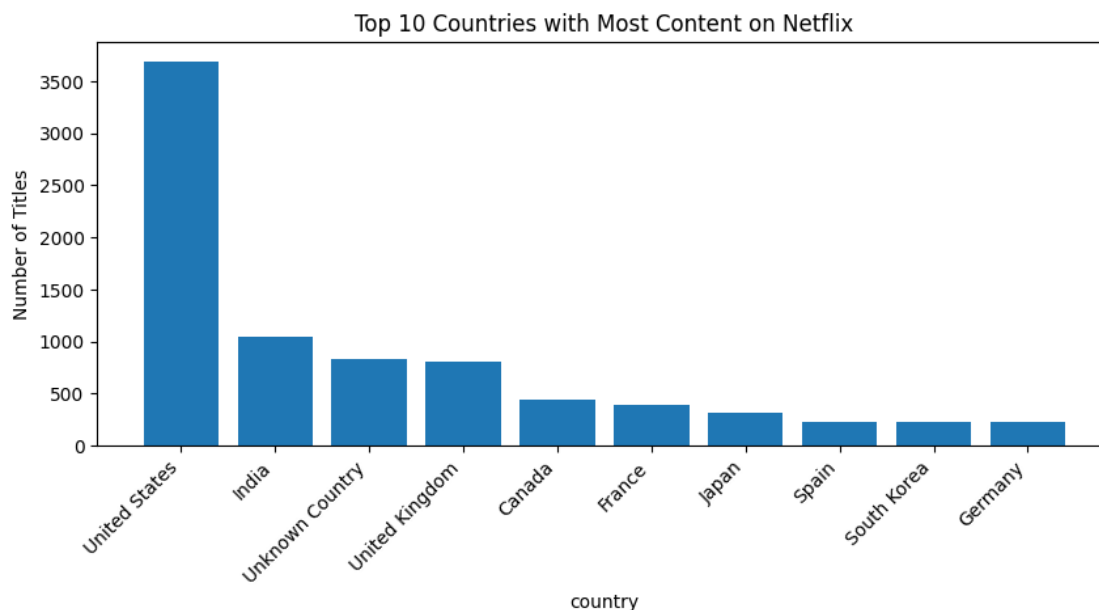
```
[969]: df_country=df.assign(country=df['country'].str.split(',')).explode('country')
df_country['country'] = df_country['country'].str.strip()
df_country['country'].value_counts()
country_counts = df_country['country'].value_counts().reset_index()
country_counts.columns = ['country', 'count']
country_counts.head(10)
```

```
[969]:
```

	country	count
0	United States	3690
1	India	1046
2	Unknown Country	831
3	United Kingdom	806
4	Canada	445
5	France	393
6	Japan	318
7	Spain	232

8	South Korea	231
9	Germany	226

```
[1005]: plt.figure(figsize=(10,4))
plt.bar(country_counts['country'][:10], country_counts['count'][:10])
plt.xticks(rotation=45, ha='right')
plt.title("Top 10 Countries with Most Content on Netflix")
plt.xlabel("country")
plt.ylabel("Number of Titles")
plt.show()
```



Insights: The maximum titles are from United States followed by India, United Kingdom, Canada, France, Japan. The titles procued by Germany is low.

### Insights:

1. The director field has the highest missing data, which may affect analysis on content creators.
2. Movies are more than TV Shows in this dataset. Customers watch more movies than TV shows overall. However from 2015 onwards Netflix significantly increased focus on TV Shows (peaking in 2017-2020).
3. These movies(Prelude to War, The Battle of Midway) are the two old movies released in year 1942. This(Pioneers: First Women Filmmakers) TV Show is too old and is released in year 1925.
4. Customers see more Movies than TV Shows. Steady growth until 2015, followed by a huge spike in movies (2017-2018) and TV Shows (2014-2018). Decline in releases post-2019, likely due to COVID-19 pandemic.
5. Peak month for adding content: July (234 TV Shows); lowest: February.

6. Most content is rated TV-MA (Mature audiences). Indicates Netflix focuses heavily on adult/mature content compared to family or children's shows.
7. International Movies are the most dominant genre, followed by Dramas and Comedies. Customers prefer International Movies over others. Thrillers, Documentaries, and Anime/K-Dramas also show significant traction.
8. Top producing countries: United States, India, United Kingdom, Japan, South Korea, Canada, France. India: Mostly recent movies (median ~2019), no significant outliers. Egypt: Older movies (median ~2010). Germany: Low number of titles compared to other regions. Highest movies watched by customers: United States.
9. Missing director names appear most frequently (many movies without known director info). Rajiv Chilaka stands out as one of the most frequent directors.
10. Netflix adds movies with famous/popular actors more frequently, suggesting celebrity-driven content strategy. Directors are more common in movies than TV shows, while TV shows are more actor-driven.

#### Recommendations:

1. Expand International Content i.e., continue prioritizing International Movies which has highest genre already.
2. Explore underrepresented regions (Africa, Middle East, Eastern Europe) to expand global audience.
3. Balance Movies & TV Shows since movies dominate overall, but TV Shows drive long-term engagement.
4. Invest in franchise-based shows (similar to K-dramas or Spanish series like La Casa de Papel).
5. Improve Data Completeness because missing values in Director, Cast, and Country limit analysis and recommendations.
6. Observe on Rating Trends since TV-MA content dominates, Netflix should keep promoting mature genres. However, adding family-friendly and kids content can attract new demographics. Content addition is highest in July — leverage this as prime release season with marketing campaigns.
7. Fill weaker months (like February) with exclusive originals or event-based releases.
8. Strengthen Local based Genres. Leverage popular actors for global reach, but also promote new directors to create fresh content. Build brand loyalty by associating directors (like Rajiv Chilaka) with Netflix Originals.