NETFLIX EDA Project

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import os
import warnings as wr
wr.filterwarnings('ignore')

os.chdir("C:\\Users\\hp\\Downloads\\Netflix data")
df =pd.read_csv("netflix_titles.csv")
df
```

```
     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
...      ...      ...                    ...              ...
8802   s8803    Movie                 Zodiac    David Fincher
8803   s8804  TV Show            Zombie Dumb              NaN
8804   s8805    Movie             Zombieland  Ruben Fleischer
8805   s8806    Movie                   Zoom     Peter Hewitt
8806   s8807    Movie                 Zubaan      Mozez Singh

                                                   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

...                                                 ...            ...

8802  Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...  United States

8803                                                NaN            NaN

8804  Jesse Eisenberg, Woody Harrelson, Emma Stone, ...  United States

8805  Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...  United States

8806  Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...          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
...                  ...           ...    ...        ...
8802   November 20, 2019          2007      R    158 min
8803        July 1, 2019          2018  TV-Y7  2 Seasons
8804    November 1, 2019          2009      R     88 min
8805    January 11, 2020          2006     PG     88 min
8806       March 2, 2019          2015  TV-14    111 min

                                                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 ...
...                                                 ...
8802                      Cult Movies, Dramas, Thrillers
8803            Kids' TV, Korean TV Shows, TV Comedies
8804                           Comedies, Horror Movies
8805                Children & Family Movies, Comedies
8806     Dramas, International Movies, Music & Musicals

                                              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...
...                                                 ...
8802  A political cartoonist, a crime reporter and a...
8803  While living alone in a spooky town, a young g...
8804  Looking to survive in a world taken over by zo...
8805  Dragged from civilian life, a former superhero...
8806  A scrappy but poor boy worms his way into a ty...

[8807 rows x 12 columns]
```

```
df.shape
```

```
(8807, 12)
```

```
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
```

```
df.describe()
```

```
       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
```

```
df.columns.tolist()
```

```
['show_id',
 'type',
 'title',
 'director',
 'cast',
 'country',
 'date_added',
 'release_year',
 'rating',
 'duration',
 'listed_in',
 'description']
```

```
df.isnull().sum()
```

```
show_id            0
type               0
title              0
director        2634
```

```
cast            825
country         831
date_added       10
release_year      0
rating            4
duration          3
listed_in         0
description       0
dtype: int64
```

```
# Filling Null values with "Unknown"
df.director.fillna(value='Unknown' ,inplace=True)
df.director
```

```
0       Kirsten Johnson
1               Unknown
2       Julien Leclercq
3               Unknown
4               Unknown
             ...
8802      David Fincher
8803            Unknown
8804    Ruben Fleischer
8805       Peter Hewitt
8806        Mozez Singh
Name: director, Length: 8807, dtype: object
```

```
df.cast.fillna(value='Unknown' , inplace= True)
df.cast
```

```
0                                                    Unknown
1       Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...
2       Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...
3                                                    Unknown
4       Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...
                              ...
8802    Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...
8803                                                 Unknown
8804    Jesse Eisenberg, Woody Harrelson, Emma Stone, ...
8805    Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...
8806    Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...
Name: cast, Length: 8807, dtype: object
```

```
df.country.fillna(value='Unknown' , inplace= True)
df.country
```

```
0       United States
1        South Africa
2             Unknown
3             Unknown
4               India
```

```
            ...
8802    United States
8803          Unknown
8804    United States
8805    United States
8806            India
Name: country, Length: 8807, dtype: object
```

```
df.date_added.fillna(value='unknown' , inplace= True)
df.date_added
```

```
0       September 25, 2021
1       September 24, 2021
2       September 24, 2021
3       September 24, 2021
4       September 24, 2021
               ...
8802     November 20, 2019
8803          July 1, 2019
8804      November 1, 2019
8805      January 11, 2020
8806         March 2, 2019
Name: date_added, Length: 8807, dtype: object
```

```
df.head()
```

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

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

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

                                        listed_in  \
0                                   Documentaries
1    International TV Shows, TV Dramas, TV Mysteries
```

```
2  Crime TV Shows, International TV Shows, TV Act...
3                            Docuseries, Reality TV
4  International TV Shows, Romantic TV Shows, TV ...

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

```
df.isnull().sum()
```

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

```
df.dropna(inplace=True)
```

```
df.isnull().sum()
```

```
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
```

```
# count of ratings
df.rating.value_counts()
```

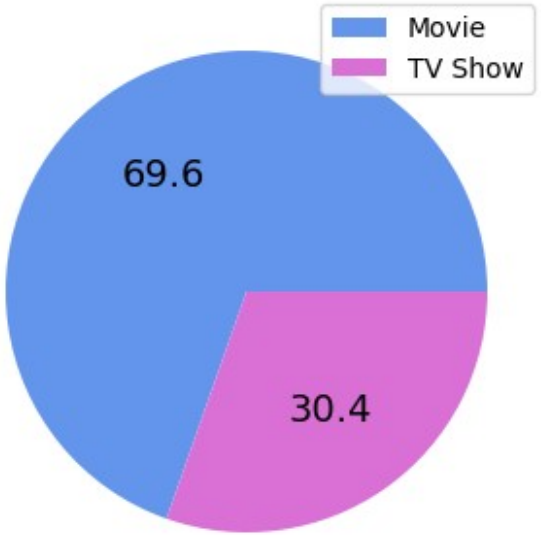
```
rating
TV-MA       3207
TV-14       2160
```

```
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
NC-17           3
UR              3
Name: count, dtype: int64
```

```
# types
df.type.value_counts().index
```

```
Index(['Movie', 'TV Show'], dtype='object', name='type')
```

```
# pie chart to see the % of movies and tv shows on Netflix
plt.figure(figsize=(4,5))
plt.pie(df.type.value_counts() , labels=df.type.value_counts().index ,
labeldistance=None , autopct= "%.1f",textprops={'fontsize':
14},colors=['cornflowerblue' , 'orchid'])
plt.legend()
plt.show()
```



Result: The movie content on Netflix is 2.33 times than TV shows

```
df.director.value_counts()
```

```
director
Unknown                                2631
Rajiv Chilaka                            19
Raúl Campos, Jan Suter                   18
Suhas Kadav                              16
Marcus Raboy                             16
                                        ...
Raymie Muzquiz, Stu Livingston            1
Joe Menendez                              1
Eric Bross                                1
Will Eisenberg                            1
Mozez Singh                               1
Name: count, Length: 4527, dtype: int64
```

Release Year

```
last_ten_yrs=df[['type' , 'release_year']]
last_ten_yrs=last_ten_yrs[last_ten_yrs["release_year"] >=2012]
last_ten_yrs
```

```
         type  release_year
0       Movie          2020
1     TV Show          2021
2     TV Show          2021
3     TV Show          2021
4     TV Show          2021
...       ...           ...
8798    Movie          2014
8800  TV Show          2012
8801    Movie          2015
8803  TV Show          2018
8806    Movie          2015

[7087 rows x 2 columns]
```

```
last_ten_yrs.groupby('release_year')['type'].value_counts()
```

```
release_year  type
2012          Movie      173
              TV Show     64
2013          Movie      225
              TV Show     62
2014          Movie      264
              TV Show     88
2015          Movie      396
              TV Show    161
2016          Movie      658
              TV Show    244
2017          Movie      765
              TV Show    265
```

```
2018          Movie      767
              TV Show    380
2019          Movie      633
              TV Show    397
2020          Movie      517
              TV Show    436
2021          TV Show    315
              Movie      277
Name: count, dtype: int64
```

```
last_ten_yrs_df=last_ten_yrs.groupby('release_year')
['type'].size().reset_index()
last_ten_yrs_df=pd.DataFrame(last_ten_yrs_df)
last_ten_yrs_df
```

	release_year	type
0	2012	237
1	2013	287
2	2014	352
3	2015	557
4	2016	902
5	2017	1030
6	2018	1147
7	2019	1030
8	2020	953
9	2021	592

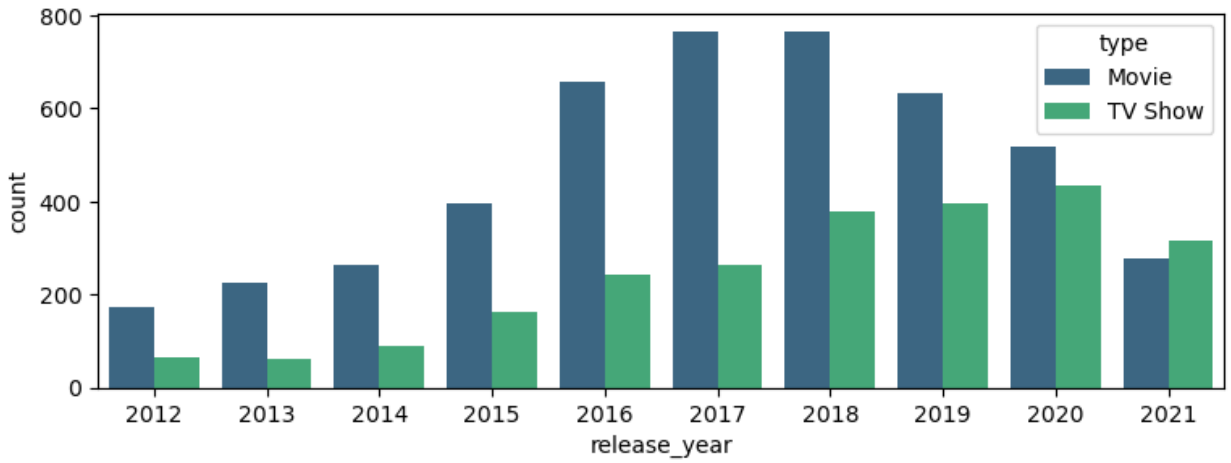
```
last_ten_yrs_df.rename(columns={"type" : "Total
Content"},inplace=True)
```

```
last_ten_yrs_df
```

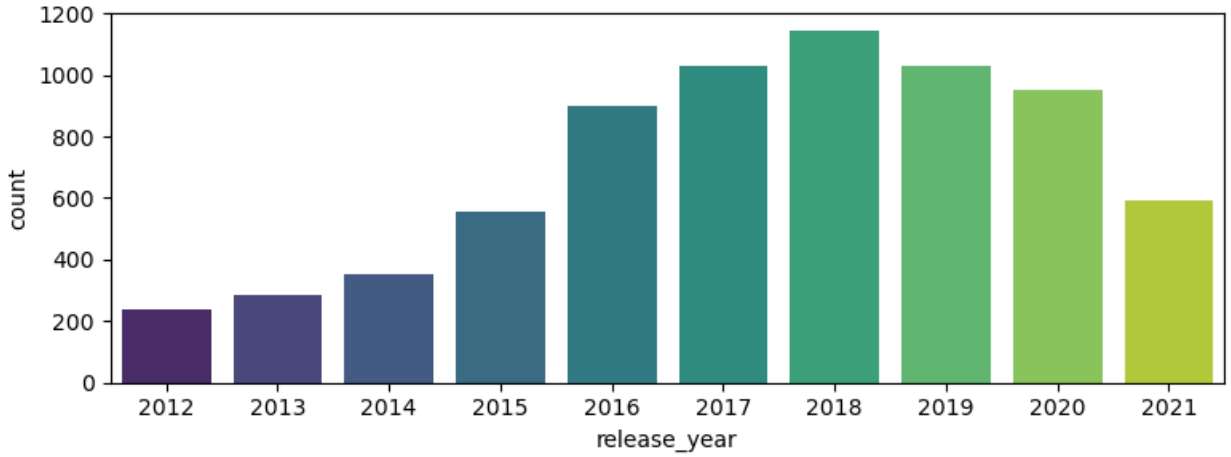
	release_year	Total Content
0	2012	237
1	2013	287
2	2014	352
3	2015	557
4	2016	902
5	2017	1030
6	2018	1147
7	2019	1030
8	2020	953
9	2021	592

```
plt.figure(figsize=(9,3))
sns.countplot(x="release_year" , data=last_ten_yrs , hue= "type",
palette= "viridis" )
```

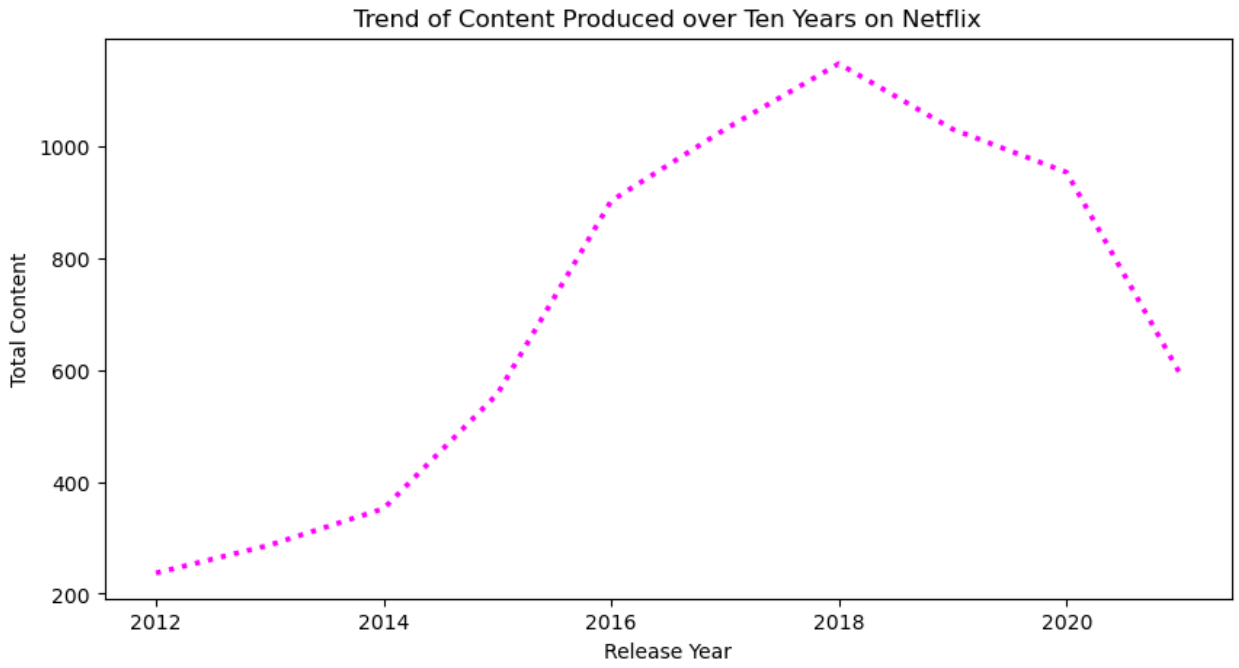
```
<Axes: xlabel='release_year', ylabel='count'>
```




```
plt.figure(figsize=(9,3))
sns.countplot(x="release_year" , data=last_ten_yrs , palette=
"viridis" )

<Axes: xlabel='release_year', ylabel='count'>
```




```
plt.figure(figsize=(10,5 ))
ax = sns.lineplot(x="release_year",y="Total
Content",data=last_ten_yrs_df, linewidth=2.5 , color='#FF00FF' ,
linestyle='dotted')
ax.set(xlabel='Release Year', ylabel='Total Content', title='Trend of
Content Produced over Ten Years on Netflix')
plt.show()
```



Result : When the last ten years are examined, it is seen that the most popular year of Netflix content is 2018. After 2018, the popularity of netflix content is decreasing.

Ratings Counts by Movies/TV Show

```
df.rating.unique()
```

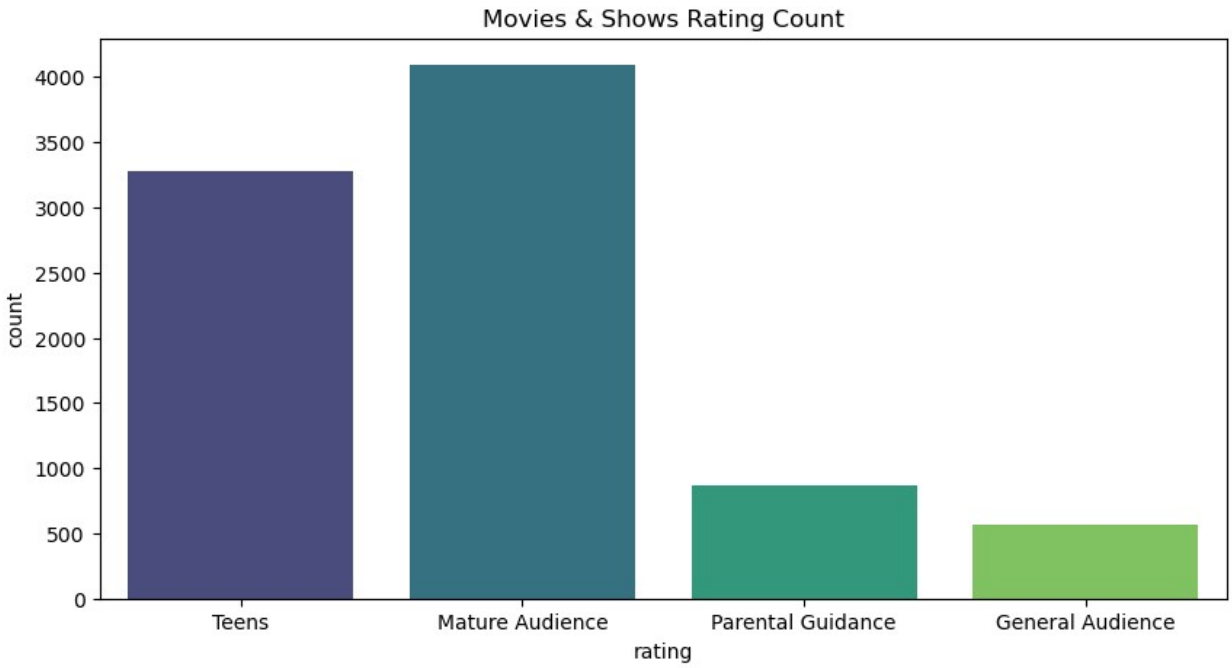
```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
       'TV-G', 'G', 'NC-17', 'NR', 'TV-Y7-FV', 'UR'], dtype=object)
```

```
new_categories = {
    'PG-13': 'Teens',
    'TV-MA': 'Mature Audience',
    'PG': 'Teens',
    'TV-14': 'Teens',
    'TV-PG': 'Parental Guidance',
    'TV-Y': 'General Audience',
    'TV-Y7': 'Teens',
    'TV-Y7-FV': 'Teens',
    'R': 'Mature Audience',
    'TV-Y': 'General Audience',
    'NR': 'Mature Audience',
    'PG-13': 'Teens',
    'TV-G': 'General Audience',
    'G': 'General Audience',
    'UR': 'Mature Audience',
    'NC-17': 'Mature Audience'
}
```

```
df["rating"] = df['rating'].replace(new_categories)
df.head()
```

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

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

           date_added  release_year           rating   duration  \
0  September 25, 2021          2020            Teens     90 min
1  September 24, 2021          2021  Mature Audience  2 Seasons
2  September 24, 2021          2021  Mature Audience   1 Season
3  September 24, 2021          2021  Mature Audience   1 Season
4  September 24, 2021          2021  Mature Audience  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...
```

```
# MOST CONTENT ON NETFLIX
plt.figure(figsize=(10,5))
sns.countplot(x="rating" , data= df , palette="viridis")
plt.title("Movies & Shows Rating Count")
plt.show()
```



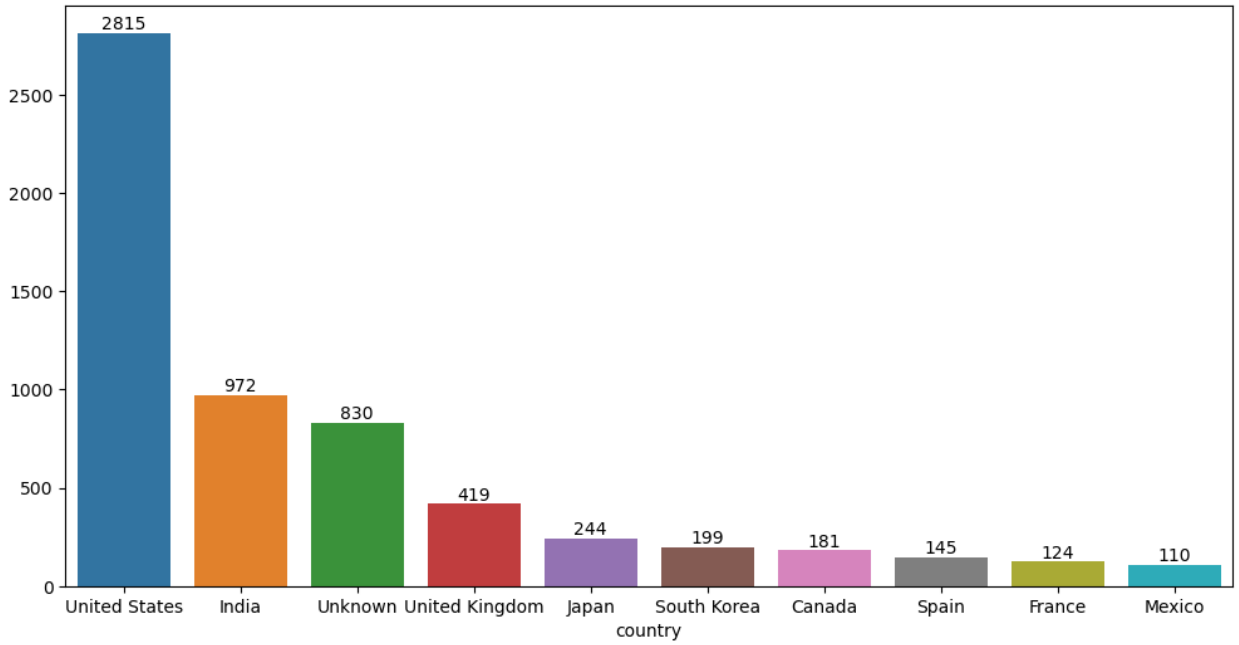
Result: It has been observed that Netflix has the most content for "Mature Audience".

Top 10 Countries where Netflix is popular

```
Top_10_country=df.country.value_counts().head(10)
Top_10_country=pd.DataFrame(Top_10_country)
Top_10_country
```

country	count
United States	2815
India	972
Unknown	830
United Kingdom	419
Japan	244
South Korea	199
Canada	181
Spain	145
France	124
Mexico	110

```
plt.figure(figsize=(12,6))
ax=sns.barplot(x=df.country.value_counts()[:10].index ,
y=df.country.value_counts()[:10].values )
ax.set_xticklabels(df.country.value_counts()[:10].index )
for i in ax.containers:
    ax.bar_label(i);
```



Result: The most content is produced in the "USA" and the second country is "India" as follows.

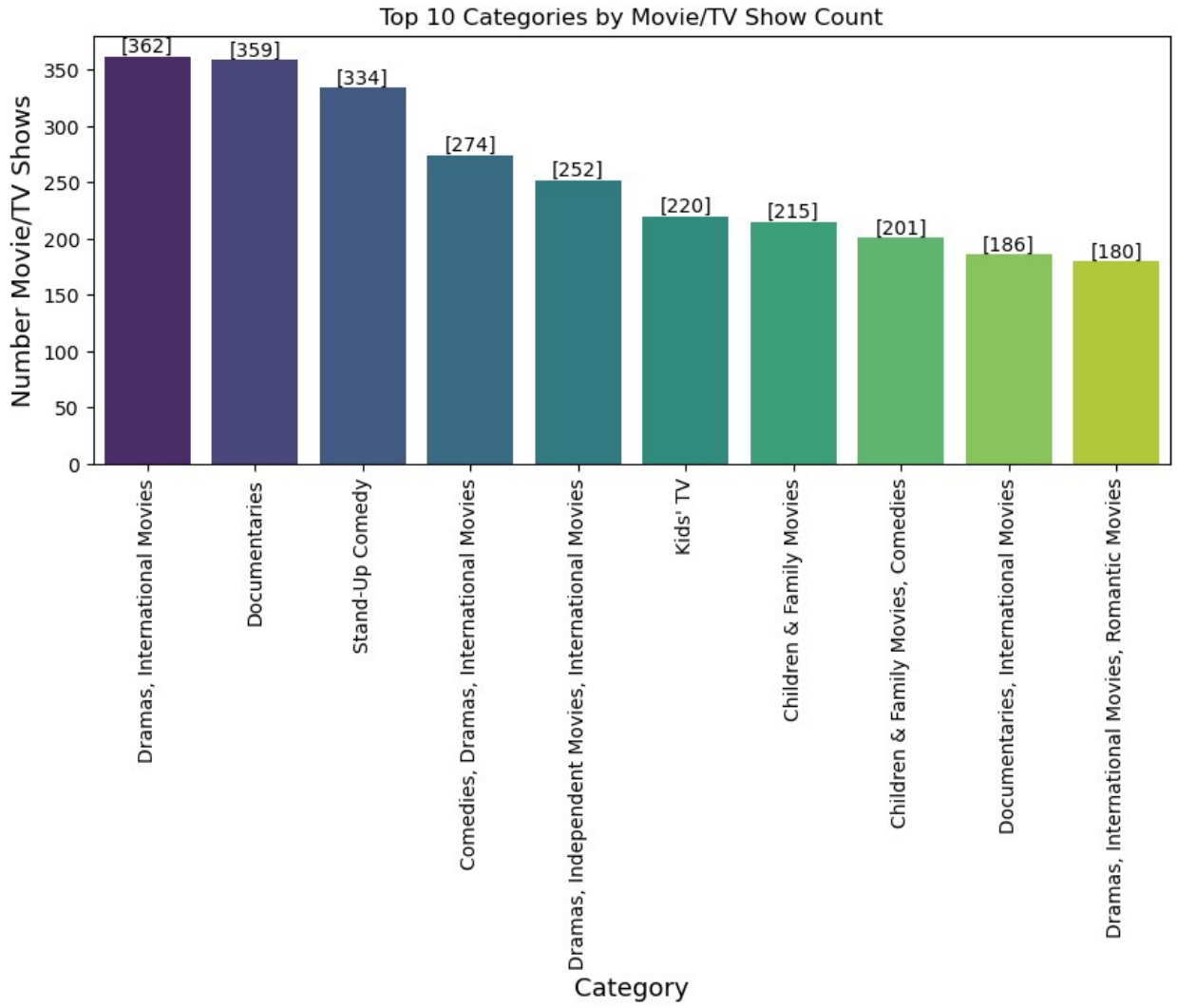
Top 10 Categories by Movie/TV Show Count

```
df.rename(columns={"listed_in" : "categories"},inplace=True)

top_10_categories=df.categories.value_counts().head(10)
top_10_categories=pd.DataFrame(top_10_categories)
top_10_categories
```

categories	count
Dramas, International Movies	362
Documentaries	359
Stand-Up Comedy	334
Comedies, Dramas, International Movies	274
Dramas, Independent Movies, International Movies	252
Kids' TV	220
Children & Family Movies	215
Children & Family Movies, Comedies	201
Documentaries, International Movies	186
Dramas, International Movies, Romantic Movies	180

```
plt.figure(figsize=(10,4))
bar_plot=sns.barplot(x=df.categories.value_counts()[:10].index ,
                     y=df.categories.value_counts()[:10].values ,
                     palette="viridis" )
ax.set_xticklabels(df.categories.value_counts()[:10].index )
plt.title('Top 10 Categories by Movie/TV Show Count')
```

```
plt.xlabel('Category', fontsize=13)
plt.ylabel('Number Movie/TV Shows',fontsize=13)
plt.xticks(rotation=90)
# Adding count values on top of each bar
for index, value in enumerate(top_10_categories.values):
    bar_plot.text(index, value, str(value), ha='center', va='bottom')

plt.show()
```



Result: The bar chart shows that " Dramas,International Movies" is the most dominant category, followed by "Documentries."