```
#import  needed libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
#load the data
netflix_df=pd.read_csv("/content/netflix_titles.csv")
print(netflix_df)
```

```
...      ...      ...                    ...              ...
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...  
```

```
#read first 5 rows
netflix_df.head()
```

	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...
					Sami							

Next steps: Generate code with `netflix_df` View recommended plots New interactive sheet

```
#read last 5 rows
netflix_df.tail()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
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...
					Jesse							

```
#check missing & null values
netflix_df.isnull().sum()
```

	0
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

```
netflix_df.isnull().any().any()
```

```
True
```

```
#cleaning the missing and null values
#fill the missing values
netflix_df['rating'].fillna(value='TV-MA',inplace=True)
netflix_df.country.fillna('United States',inplace=True)
netflix_df.dropna(inplace=True)
#check missing values
netflix_df.isnull().sum()
```

	0
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

```
#get info about the dataset
netflix_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 5697 entries, 2 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       5697 non-null   object
 1   type          5697 non-null   object
 2   title         5697 non-null   object
 3   director      5697 non-null   object
 4   cast          5697 non-null   object
 5   country       5697 non-null   object
 6   date_added    5697 non-null   object
 7   release_year  5697 non-null   int64 
 8   rating        5697 non-null   object
 9   duration      5697 non-null   object
 10  listed_in     5697 non-null   object
```

```
      11  description   5697 non-null   object
     dtypes: int64(1), object(11)
     memory usage: 578.6+ KB
```

```
#description only release_year
netflix_df['release_year'].describe()
```

	release_year
count	5697.000000
mean	2012.978936
std	9.564384
min	1942.000000
25%	2012.000000
50%	2016.000000
75%	2018.000000
max	2021.000000

dtype: float64

```
#columns in the dataset
netflix_df.columns
```

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

```
#shape of the dataset
netflix_df.shape
```

```
(5697, 12)
```

```
#how many years of data(1966-2021)(release_year)
start_year = 1966
end_year = 2021
#calculating then num of years
num_years = end_year - start_year + 1
print(f"The number of years in the dataset from {start_year} to {end_year} is: {num_years}")
```

```
The number of years in the dataset from 1966 to 2021 is: 56
```

```
#check null values for all columns
netflix_df.isnull().sum()
```

	0
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

```
netflix_df['rating'].fillna( 'TV-MA',inplace  = True)
netflix_df['rating']
```

	rating
2	TV-MA
5	TV-MA
6	PG
7	TV-MA
8	TV-14
...	...
8801	TV-MA
8802	R
8804	R
8805	PG
8806	TV-14

5697 rows × 1 columns

dtype: object

```
netflix_df['country'].fillna ("United States",inplace=True)
netflix_df['country']
```

	country
2	United States
5	United States
6	United States
7	United States, Ghana, Burkina Faso, United Kin...
8	United Kingdom
...	...
8801	United Arab Emirates, Jordan
8802	United States
8804	United States
8805	United States
8806	India

5697 rows × 1 columns

dtype: object

```
netflix_df.rename(columns={'listed_in': 'genre'},inplace=True)
print(netflix_df.columns)
```

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

```
netflix_df['country'].describe()
```

	country
count	5697
unique	604
top	United States
freq	2210

dtype: object

```
#get value counts for the type column
netflix=pd.DataFrame(netflix_df['type'].value_counts())
print(netflix)
```

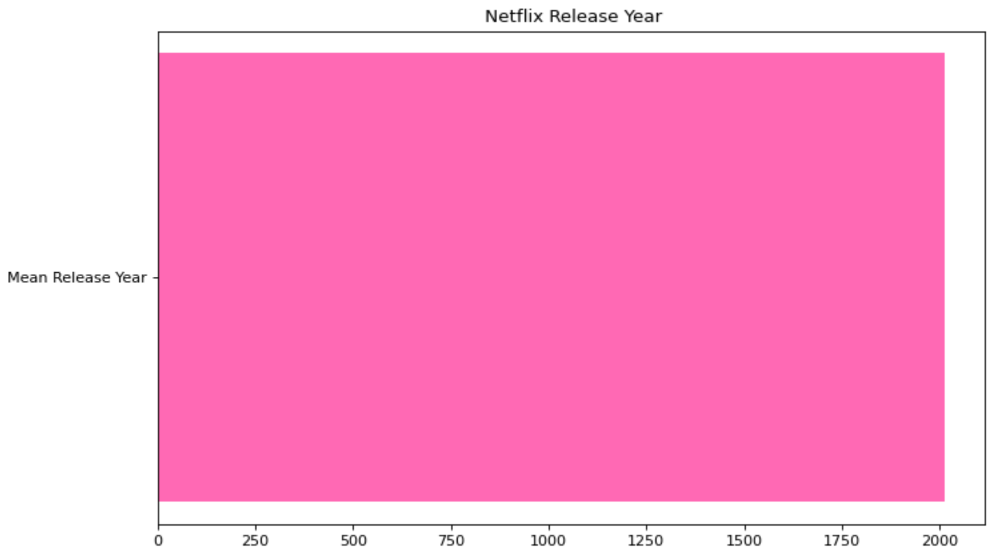
```
          count
type
Movie      5519
TV Show     178
```

```
VISUALIZATION
```

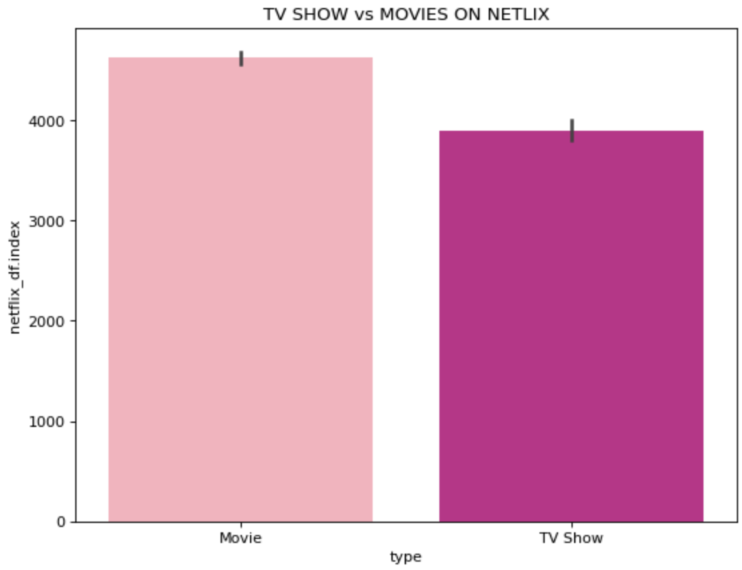
```
#Visualizaion
#calculating mean  of release_year and show horizontal bargraph
netflix_df["release_year"].mean()
```

```
2012.978936282254
```

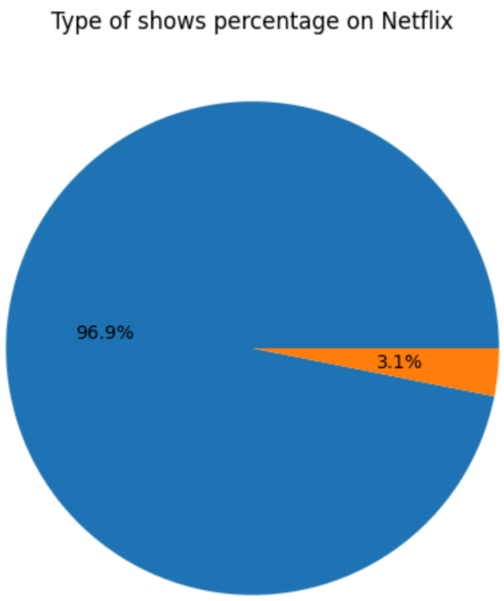
```
#horizontal bargraph
plt.figure(figsize=(10,6),dpi=80)
mean_release_year=netflix_df['release_year'].mean()
plt.barh(['Mean Release Year'],[mean_release_year],color='hotpink')
#add title
plt.title(' Netflix Release Year')
plt.show()
```



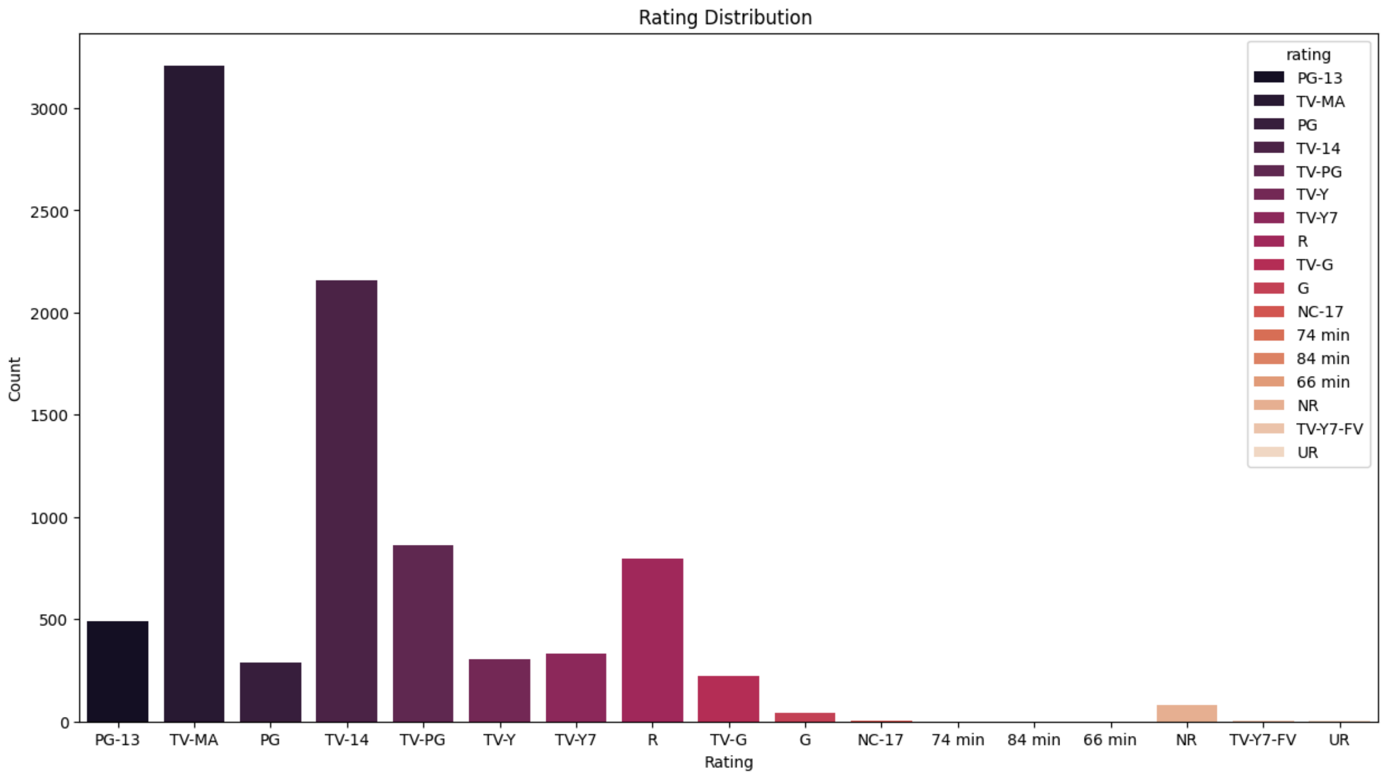
```
#Tv Show vs Movie Barchart
plt.figure(figsize=(8,6),dpi=80)
sns.barplot(x=netflix_df['type'],y=netflix_df.index,palette='RdPu',hue=netflix_df['type'])
#add labels xaxis yaxis
plt.xlabel('type')
plt.ylabel('netflix_df.index')
#add tittle
plt.title('TV SHOW vs MOVIES ON NETLIX')
plt.show()
```



```
#Type of shows percentage on TV SHOWS & MOVIES
fig = plt.figure(figsize=(10,6))
#plot pie
plt.pie(netflix_df['type'].value_counts(),autopct='%1.1f%%',)
plt.title("Type of shows percentage on Netflix")
plt.show()
```



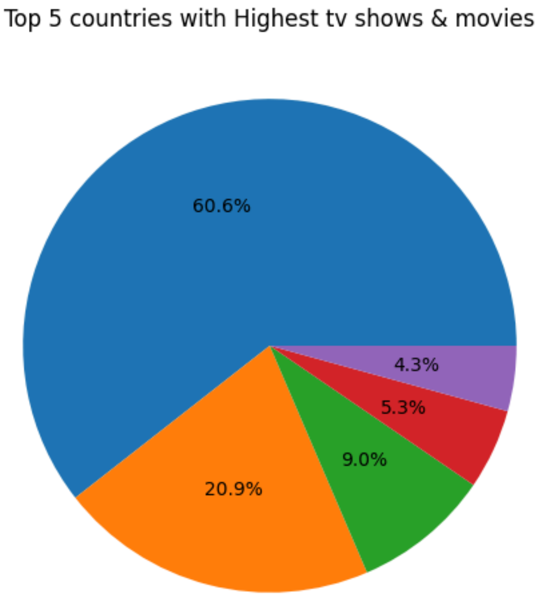
```
#Netflix  Rating  Distribution
plt.figure(figsize=(15,8))
#plot countplot
sns.countplot(data=netflix_df,x='rating',palette='rocket',hue='rating')
#add labels
plt.xlabel('Rating')
plt.ylabel('Count')
#add title
plt.title('Rating Distribution ')
#plot show
plt.show()
```



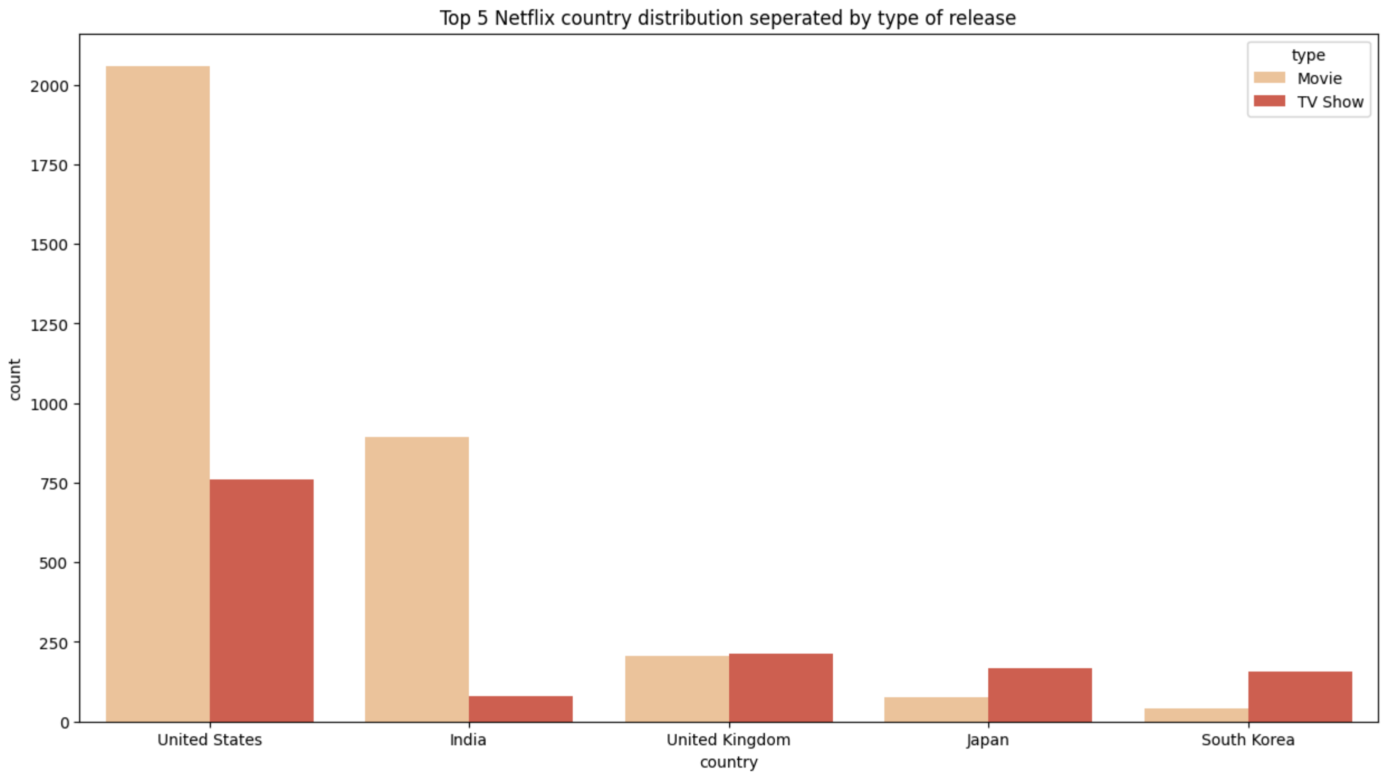
```
netflix_df=pd.DataFrame (netflix_df['country'].value_counts())
print(netflix_df)
```

```
                                               count
country
United States                                   2210
India                                            875
United Kingdom                                   183
Canada                                           107
Spain                                             91
...                                              ...
Uruguay, Guatemala                                 1
Romania, Bulgaria, Hungary                         1
Philippines, United States                         1
India, United Kingdom, Canada, United States       1
United Arab Emirates, Jordan                       1

[604 rows x 1 columns]
```

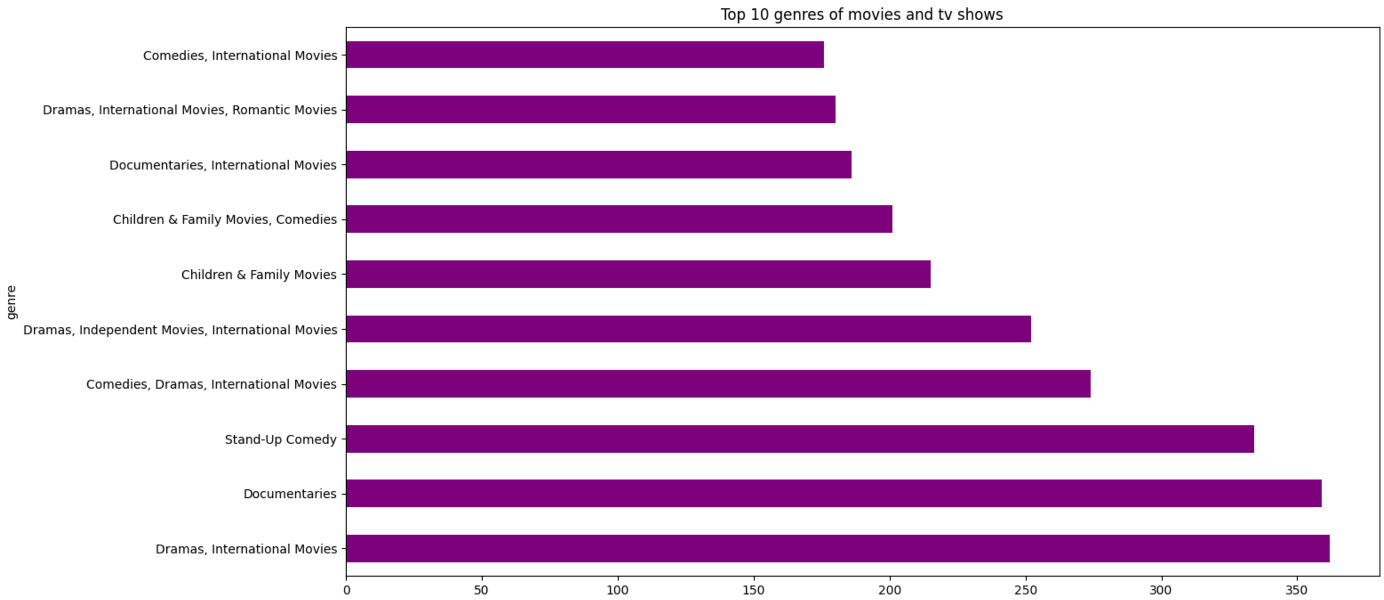
```
#top 5 countries in highest Tv shows/Movies
import matplotlib.pyplot as plt
import seaborn as sns
fig = plt.figure(figsize=(10,6))
#plot pie
plt.pie(netflix_df['country'].value_counts().nlargest(n=5),autopct='%1.1f%%')
#add title
plt.title('Top 5 countries with Highest tv shows & movies')
plt.show()
```



```
#top 5 country distribution seperated by type of release bar graph
plt.figure(figsize = (15,8))

#plot countplot
sns.countplot(x='country', data=netflix_df, hue='type', order=netflix_df.country.value_counts().iloc[:5].index, palette="OrRd")

plt.title('Top 5 Netflix country distribution seperated by type of release')
plt.show()
```

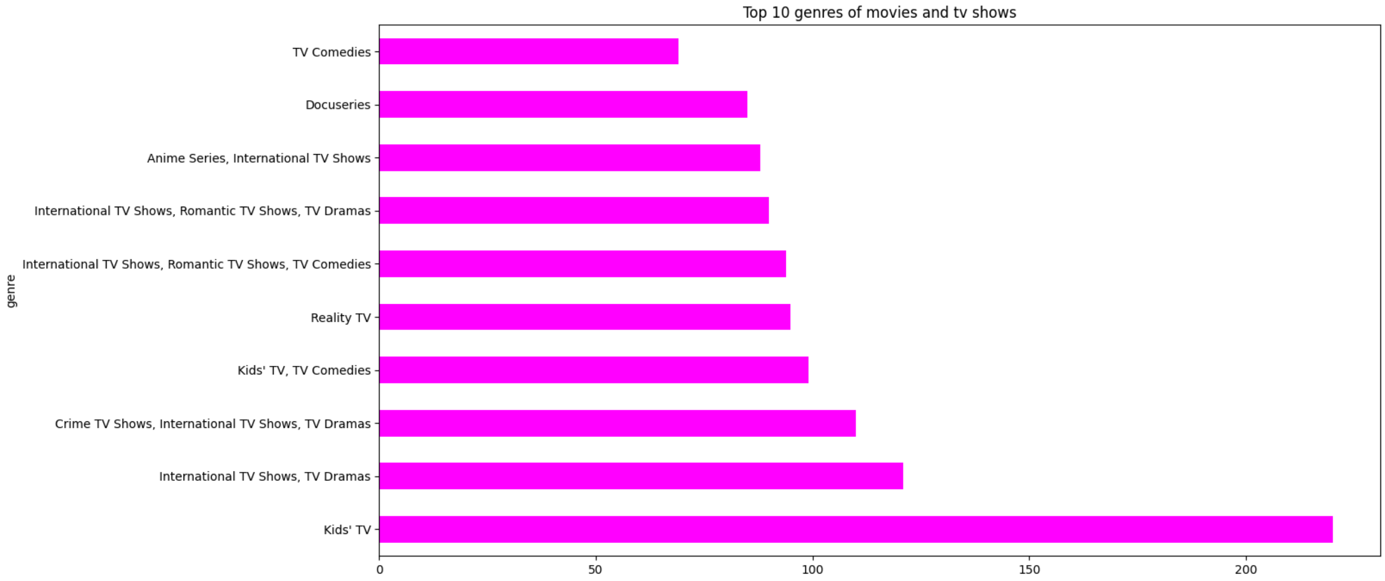


```
#Top 10 genres of movies and tv shows
plt.figure(figsize=(15,8))
netflix_df[netflix_df['type']=='Movie']['genre'].value_counts()[:10].plot(kind='barh', color='purple')
#add title
plt.title('Top 10 genres of movies and tv shows')
#plot show
plt.show()
```

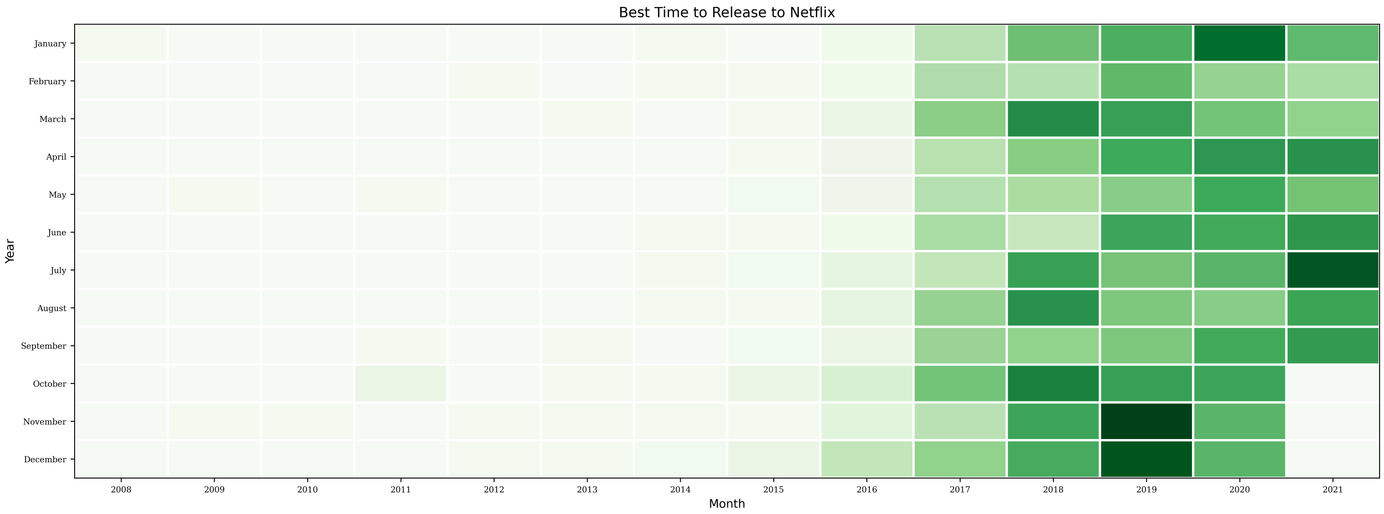


```
#Top 10 genres of tv shows
plt.figure(figsize=(15,8))
netflix_df[netflix_df["type"]=="TV Show"]["genre"].value_counts()[:10].plot(kind='barh', color='magenta')
#add title
```

```
plt.title('Top 10 genres of movies and tv shows')
#plot show
plt.show()
```



```
#9.Best time to release  Tv Shows and  Movies

netflix_df = netflix_df[['date_added']].dropna()
netflix_df['year'] = netflix_df['date_added'].apply(lambda x : x.split(', ')[-1])
netflix_df['month'] = netflix_df['date_added'].apply(lambda x : x.lstrip().split(' ')[0])


month_order = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December'][::-1
df = netflix_df.groupby('year')['month'].value_counts().unstack().fillna(0)[month_order].T


plt.figure(figsize=(20, 7), dpi=500)
plt.pcolor(df, cmap='Greens', edgecolors='white', linewidths=2)

plt.xticks(np.arange(0.5, len(df.columns), 1), df.columns, fontsize=7, fontfamily='serif')
plt.yticks(np.arange(0.5, len(df.index), 1), df.index, fontsize=7, fontfamily='serif')
#add labels
plt.xlabel('Month', fontsize=10)
plt.ylabel('Year', fontsize=10)
#add title
plt.title('Best Time to Release to Netflix', fontsize=12, position=(0.50, 1.0+0.02))


#plot show
plt.show()
```


Best Time to Release to Netflix
Year
January
February
March
April
May
June
July
August
September
October
November
December
Month