NETFLIX LOGO

NETFLIX DATA ANALYSIS

IMPORT THE LIBRARIES

In [1]:

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

LOAD THE DATA

In [2]:

```

data=pd.read_csv('C:\\Users\\Hp\\Desktop\\NETFLIX DAP\\netflix_dataset.csv')
data
```

Out[2]:

	show_id	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	type
0	81193313	Chocolate	NaN	Ha Ji-won, Yoon Kye-sang, Jang Seung-jo, Kang ...	South Korea	November 30, 2019	2019	TV-14	1 Season	International TV Shows, Korean TV Shows, Roman...	Brought together by meaningful meals in the pa...	TV Show
1	81197050	Guatemala: Heart of the Mayan World	Luis Ara, Ignacio Jaunsolo	Christian Morales	NaN	November 30, 2019	2019	TV-G	67 min	Documentaries, International Movies	From Sierra de las Minas to Esquipulas, explor...	Movie
2	81213894	The Zoya Factor	Abhishek Sharma	Sonam Kapoor, Dulquer Salmaan, Sanjay Kapoor, ...	India	November 30, 2019	2019	TV-14	135 min	Comedies, Dramas, International Movies	A goofy copywriter unwittingly convinces the I...	Movie
3	81082007	Atlantics	Mati Diop	Mama Sane, Amadou Mbow, Ibrahima Traore, Nicol...	France, Senegal, Belgium	November 29, 2019	2019	TV-14	106 min	Dramas, Independent Movies, International Movies	Arranged to marry a rich man, young Ada is cru...	Movie
4	80213643	Chip and Potato	NaN	Abigail Oliver, Andrea Libman, Briana Buckmast...	Canada, United Kingdom	NaN	2019	TV-Y	2 Seasons	Kids' TV	Lovable pug Chip starts kindergarten, makes ne...	TV Show
...	...	...	...	...	...	...	...	...	...	...	...	...
5832	70141644	Mad Ron's Prevues from Hell	Jim Monaco	Nick Pawlow, Jordu Schell, Jay Kushwara, Micha...	United States	November 1, 2010	1987	NR	84 min	Cult Movies, Horror Movies	This collection cherry-picks trailers, forgott...	Movie
5833	70127998	Splatter	Joe Dante	Corey Feldman, Tony Todd, Tara Leigh, Erin Way...	United States	November 18, 2009	2009	TV-14	29 min	Horror Movies	After committing suicide, a washed-up rocker r...	Movie
5834	70084180	Just Another Love Story	Ole Bornedal	Anders W. Berthelsen, Rebecka Hemse, Nikolaj L...	Denmark	May 5, 2009	2007	NR	104 min	Dramas, International Movies	When he causes a car accident that leaves a yo...	Movie
5835	70157452	Dinner for Five	NaN	NaN	United States	February 4, 2008	2007	TV-MA	1 Season	Stand-Up Comedy & Talk Shows	In each episode, four celebrities join host Jo...	TV Show
5836	70053412	To and From New York	Sorin Dan Mihalcescu	Barbara King, Shaana Diya, John Krisiukenas, Y...	United States	January 1, 2008	2006	NR	81 min	Dramas, Independent Movies, Thrillers	While covering a story in New York City, a Sea...	Movie

5837 rows × 12 columns

In [3]:

```
data.shape #used to find the dimensions of dataset(numberofrows, numberofcolumns)
```

Out[3]:

```
(5837, 12)
```

In [4]:

```
data.head() #it can print first 5 rows by default
```

Out[4]:

	show_id	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	type
0	81193313	Chocolate	NaN	Ha Ji-won, Yoon Kye-sang, Jang Seung-jo, Kang ...	South Korea	November 30, 2019	2019	TV-14	1 Season	International TV Shows, Korean TV Shows, Roman...	Brought together by meaningful meals in the pa...	TV Show
1	81197050	Guatemala: Heart of the Mayan World	Luis Ara, Ignacio Jaunsolo	Christian Morales	NaN	November 30, 2019	2019	TV-G	67 min	Documentaries, International Movies	From Sierra de las Minas to Esquipulas, explor...	Movie
2	81213894	The Zoya Factor	Abhishek Sharma	Sonam Kapoor, Dulquer Salmaan, Sanjay Kapoor, ...	India	November 30, 2019	2019	TV-14	135 min	Comedies, Dramas, International Movies	A goofy copywriter unwittingly convinces the I...	Movie
3	81082007	Atlantics	Mati Diop	Mama Sane, Amadou Mbow, Ibrahima Traore, Nicol...	France, Senegal, Belgium	November 29, 2019	2019	TV-14	106 min	Dramas, Independent Movies, International Movies	Arranged to marry a rich man, young Ada is cru...	Movie
4	80213643	Chip and Potato	NaN	Abigail Oliver, Andrea Libman, Briana Buckmast...	Canada, United Kingdom	NaN	2019	TV-Y	2 Seasons	Kids' TV	Lovable pug Chip starts kindergarten, makes ne...	TV Show

In [5]:

```
data.tail(10) #it an print last 10 rows
```

Out[5]:

	show_id	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	type
5827	70206824	Triumph of the Heart	Richard Michaels	Mario Van Peebles, Susan Ruttan, Lane R. Davis...	United States	October 1, 2011	1991	TV-PG	93 min	Dramas, Sports Movies	This drama tells the tale of Ricky Bell, a pro...	Movie
5828	70206825	Unspeakable Acts	Linda Otto	Jill Clayburgh, Brad Davis, Sam Behrens	United States	October 1, 2011	1990	NR	95 min	Dramas	Laurie and Joseph are doctors who interview se...	Movie
5829	70206826	Victim of Beauty	Roger Young	William Devane, Jeri Ryan, Michele Abrams, Nic...	United States	October 1, 2011	1991	NR	93 min	Dramas, Thrillers	A beauty pageant winner is stalked by a serial...	Movie
5830	60003155	Joseph: King of Dreams	Rob LaDuca, Robert C. Ramirez	Ben Affleck, Mark Hamill, Richard Herd, Mauree...	United States	September 27, 2011	2000	TV-PG	75 min	Children & Family Movies, Dramas, Faith & Spir...	With his gift of dream interpretation and his ...	Movie
5831	70154110	Even the Rain	Icíar Bollaín	Luis Tosar, Gael García Bernal, Juan Carlos Ad...	Spain, Mexico, France	May 17, 2011	2010	NR	103 min	Dramas, International Movies	While making a film about the incursion of Chr...	Movie
5832	70141644	Mad Ron's Prevues from Hell	Jim Monaco	Nick Pawlow, Jordu Schell, Jay Kushwara, Micha...	United States	November 1, 2010	1987	NR	84 min	Cult Movies, Horror Movies	This collection cherry-picks trailers, forgott...	Movie
5833	70127998	Splatter	Joe Dante	Corey Feldman, Tony Todd, Tara Leigh, Erin Way...	United States	November 18, 2009	2009	TV-14	29 min	Horror Movies	After committing suicide, a washed-up rocker r...	Movie
5834	70084180	Just Another Love Story	Ole Bornedal	Anders W. Berthelsen, Rebecka Hemse, Nikolaj L...	Denmark	May 5, 2009	2007	NR	104 min	Dramas, International Movies	When he causes a car accident that leaves a yo...	Movie
5835	70157452	Dinner for Five	NaN	NaN	United States	February 4, 2008	2007	TV-MA	1 Season	Stand-Up Comedy & Talk Shows	In each episode, four celebrities join host Jo...	TV Show
5836	70053412	To and From New York	Sorin Dan Mihalcescu	Barbara King, Shaana Diya, John Krisiukenas, Y...	United States	January 1, 2008	2006	NR	81 min	Dramas, Independent Movies, Thrillers	While covering a story in New York City, a Sea...	Movie

In [6]:
```
data.info() #datatypes and summary of the metadata
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5837 entries, 0 to 5836
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       5837 non-null   int64 
 1   title         5837 non-null   object
 2   director      3936 non-null   object
 3   cast          5281 non-null   object
 4   country       5410 non-null   object
 5   date_added    5195 non-null   object
 6   release_year  5837 non-null   int64 
 7   rating        5827 non-null   object
 8   duration      5837 non-null   object
 9   listed_in     5837 non-null   object
 10  description   5837 non-null   object
 11  type          5837 non-null   object
dtypes: int64(2), object(10)
memory usage: 547.3+ KB
```

In [7]:
```
data.isnull() #used to check is there any null values in data
```

Out[7]:

	show_id	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	type
0	False	False	True	False	False	False	False	False	False	False	False	False
1	False	False	False	False	True	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	True	False	False	True	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...
5832	False	False	False	False	False	False	False	False	False	False	False	False
5833	False	False	False	False	False	False	False	False	False	False	False	False
5834	False	False	False	False	False	False	False	False	False	False	False	False
5835	False	False	True	True	False	False	False	False	False	False	False	False
5836	False	False	False	False	False	False	False	False	False	False	False	False

5837 rows × 12 columns

In [8]:
```
data.describe() #it provides a statistical summary of only numerical columns in the dataset.
```

Out[8]:

	show_id	release_year
count	5.837000e+03	5837.000000
mean	7.730079e+07	2013.688539
std	9.479777e+06	8.419088
min	2.698800e+05	1925.000000
25%	8.004520e+07	2013.000000
50%	8.016353e+07	2016.000000
75%	8.024188e+07	2018.000000
max	8.122720e+07	2020.000000

In [9]:
```
#i want to rename the column name diretor into Directors_Name
data.rename(columns={'director':'Directors'},inplace=True) #Rename the columns in the original DataFrame without creating new dataframe
```

In [10]:
```
data.columns #it display the column names in the dataframe. and column renamed director to Directors
```

Out[10]:
```
Index(['show_id', 'title', 'Directors', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description',
       'type'],
      dtype='object')
```

In [11]:
```
data.nunique()  #Counts unique values in each column
```

Out[11]:
```
show_id         5837
title           5780
Directors       3108
cast            5087
country          527
date_added      1092
release_year      71
rating            14
duration         194
listed_in        449
description     5829
type               2
dtype: int64
```

In [12]:

```
data.isnull().sum() #it shows total number of null values in each column
```

Out[12]:

```
show_id            0
title              0
Directors       1901
cast             556
country          427
date_added       642
release_year       0
rating            10
duration           0
listed_in          0
description        0
type               0
dtype: int64
```

In [13]:

```
#i want to add the missing value (null value) in the column of director
data['Directors'].fillna('not spicified', inplace=True)
data['cast'].fillna('not spicified', inplace=True)
data['country'].fillna('not spicified', inplace=True)

#Replaced all null values in the column director with the string 'not specified
```

In [14]:

```
data.dropna(subset=['date_added'],inplace=True) #This will remove rows where the date_added column has NaN values.
data.dropna(subset=['rating'],inplace=True)
```

In [15]:

```
data.isnull().sum() #after replaced the null values in the all columns is zero
```

Out[15]:

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

```
In [16]: data.duplicated() #Helps to detect duplicate rows in the dataset.
```

```
Out[16]: 0       False
         1       False
         2       False
         3       False
         5       False
                 ...
         5832    False
         5833    False
         5834    False
         5835    False
         5836    False
         Length: 5186, dtype: bool
```

```
In [17]: data.duplicated().sum() #checking total number of dupliate values
```

```
Out[17]: 0
```

```
In [18]: data['type'].unique() #it will return an array of unique values in the 'type' colums
```

```
Out[18]: array(['TV Show', 'Movie'], dtype=object)
```

```
In [19]: data['title'].unique() #return an array of unique values in the "title" column.
```

```
Out[19]: array(['Chocolate', 'Guatemala: Heart of the Mayan World',
                'The Zoya Factor', ..., 'Just Another Love Story',
                'Dinner for Five', 'To and From New York'], dtype=object)
```

```
In [20]: data['Directors'].unique() #It checks the Directors column and returns a list of unique values in that column, meaning no duplicates.
```

```
Out[20]: array(['not spicified', 'Luis Ara, Ignacio Jaunsolo', 'Abhishek Sharma',
                ..., 'Jim Monaco', 'Ole Bornedal', 'Sorin Dan Mihalcescu'],
               dtype=object)
```

```
In [21]: data['cast'].unique() #return an array of unique values in the "cast" column.
```

```
Out[21]: array(['Ha Ji-won, Yoon Kye-sang, Jang Seung-jo, Kang Bu-ja, Lee Jae-ryong, Min Jin-woong, Kim Won-hae, Yoo Teo',
                'Christian Morales',
                'Sonam Kapoor, Dulquer Salmaan, Sanjay Kapoor, Sikander Kher, Angad Bedi, Koel Purie, Pooja Bhamrah, Manu Rishi Chadha',
                ...,
                'Corey Feldman, Tony Todd, Tara Leigh, Erin Way, Mark Alan, Stuart Pankin',
                'Anders W. Berthelsen, Rebecka Hemse, Nikolaj Lie Kaas, Charlotte Fich, Dejan Cukic, Karsten Jansfort, Flemming Enevold, Bent Mejding, Ewa Fröling, Josephine Raahauge, Timm Vla
         dimir, Ditte Hansen',
                'Barbara King, Shaana Diya, John Krisiukenas, Yoriko Haraguchi, David Crane, Chad Thomas'],
               dtype=object)
```

```
In [22]: data['date_added'].unique() #return an array of unique values in the "date_added" column.
```

```
Out[22]: array(['November 30, 2019', 'November 29, 2019', 'November 28, 2019', ...,
                'May 5, 2009', 'February 4, 2008', 'January 1, 2008'], dtype=object)
```

```
In [23]: data['rating'].unique()
```

```
Out[23]: array(['TV-14', 'TV-G', 'TV-MA', 'TV-PG', 'R', 'TV-Y7', 'TV-Y', 'PG', 'G',
                'PG-13', 'TV-Y7-FV', 'NR', 'UR', 'NC-17'], dtype=object)
```

```
In [24]: data['release_year'].unique()
```

```
Out[24]: array([2019, 2018, 2016, 2017, 2001, 2008, 1980, 1973, 2012, 1987, 2002,
                2015, 2013, 1996, 1992, 2007, 2009, 1995, 2014, 1968, 2010, 1999,
                2011, 2000, 1974, 1997, 1993, 2006, 2005, 2003, 2004, 2020, 1998,
                1960, 1958, 1965, 1956, 1978, 1990, 1976, 1962, 1955, 1972, 1991,
                1988, 1989, 1975, 1985, 1984, 1970, 1994, 1982, 1983, 1967, 1981,
                1979, 1986, 1925, 1946, 1977, 1959, 1966, 1971, 1969, 1945, 1942,
                1944, 1947, 1943, 1954], dtype=int64)
```

```
In [25]: data['duration'].unique()
```

```
Out[25]: array(['1 Season', '67 min', '135 min', '106 min', '107 min', '81 min',
                '118 min', '110 min', '104 min', '93 min', '94 min', '124 min',
                '137 min', '134 min', '69 min', '209 min', '86 min', '24 min',
                '46 min', '117 min', '87 min', '92 min', '114 min', '121 min',
                '109 min', '96 min', '97 min', '56 min', '119 min', '138 min',
                '111 min', '88 min', '73 min', '116 min', '85 min', '102 min',
                '101 min', '28 min', '103 min', '131 min', '166 min', '105 min',
                '82 min', '84 min', '112 min', '89 min', '136 min', '129 min',
                '158 min', '78 min', '100 min', '74 min', '60 min', '143 min',
                '98 min', '54 min', '59 min', '95 min', '61 min', '123 min',
                '44 min', '68 min', '99 min', '91 min', '40 min', '90 min',
                '108 min', '200 min', '133 min', '115 min', '55 min', '153 min',
                '185 min', '127 min', '120 min', '139 min', '122 min', '36 min',
                '141 min', '65 min', '126 min', '63 min', '14 min', '20 min',
                '52 min', '83 min', '66 min', '140 min', '22 min', '62 min',
                '151 min', '154 min', '70 min', '76 min', '77 min', '45 min',
                '147 min', '58 min', '64 min', '125 min', '79 min', '163 min',
                '42 min', '38 min', '146 min', '130 min', '152 min', '182 min',
                '171 min', '80 min', '157 min', '142 min', '10 min', '128 min',
                '149 min', '113 min', '47 min', '167 min', '72 min', '145 min',
                '164 min', '177 min', '57 min', '161 min', '32 min', '53 min',
                '26 min', '48 min', '176 min', '15 min', '71 min', '12 min',
                '3 min', '30 min', '159 min', '150 min', '165 min', '148 min',
                '49 min', '168 min', '170 min', '132 min', '75 min', '162 min',
                '51 min', '50 min', '144 min', '34 min', '29 min', '312 min',
                '27 min', '35 min', '155 min', '23 min', '11 min', '205 min',
                '190 min', '25 min', '214 min', '156 min', '31 min', '196 min',
                '33 min', '203 min', '160 min', '19 min', '169 min', '173 min',
                '195 min', '41 min', '187 min', '193 min', '192 min', '224 min',
                '172 min', '18 min', '179 min', '180 min', '43 min'], dtype=object)
```

```
In [26]: data['country'].unique()
```

```
         'United States, United Kingdom, Canada', 'China, Hong Kong',
         'Canada, United States', 'Canada',
         'South Africa, United States, New Zealand, Canada',
         'United States, Canada', 'United States, Australia, Mexico',
         'United States, South Korea, Japan',
         'United States, Italy, United Kingdom, Liechtenstein',
         'Hong Kong, United States', 'Italy',
         'United States, United Kingdom, Italy',
         'United States, France, Canada, Spain', 'Mexico, United States',
         'United Kingdom, United States', 'United States, Mexico',
         'United States, United Kingdom', 'Hong Kong',
         'United States, Australia', 'Italy, Belgium, France',
         'Argentina, Spain', 'United States, United Kingdom, France',
         'Israel, United States', 'Turkey',
         'United States, United Kingdom, Germany', 'Switzerland',
         'Singapore, Malaysia', 'Canada, Luxembourg', 'Hong Kong, China',
         'United States, Spain, Germany', 'Egypt, Austria, United States',
         'Belgium, Luxembourg, France', 'Indonesia', 'United States, China',
         'Australia', 'United Kingdom, Poland, United States',
         'Spain, France', 'Australia, United States',
```

```
In [27]: data['listed_in'].unique()
```

```
Out[27]: array(['International TV Shows, Korean TV Shows, Romantic TV Shows',
         'Documentaries, International Movies',
         'Comedies, Dramas, International Movies',
         'Dramas, Independent Movies, International Movies',
         'Comedies, International Movies, Thrillers',
         'Dramas, International Movies', 'Comedies, International Movies',
         'Comedies, Independent Movies, International Movies', 'Reality TV',
         'Action & Adventure, Comedies, International Movies',
         'International TV Shows, Romantic TV Shows, TV Comedies',
         'Dramas, International Movies, Thrillers', 'Docuseries',
         'Children & Family Movies, Dramas',
         'Anime Series, International TV Shows',
         'International TV Shows, Stand-Up Comedy & Talk Shows, TV Comedies',
         'TV Comedies', 'International TV Shows, TV Comedies, TV Dramas',
         'Dramas, Thrillers', 'Action & Adventure, Dramas',
         'Comedies, International Movies, Romantic Movies',
         'Children & Family Movies, Comedies', 'Dramas, Independent Movies',
         'Dramas', 'Stand-Up Comedy', 'Children & Family Movies', 'Movies',
         'International TV Shows, Spanish-Language TV Shows, Stand-Up Comedy & Talk Shows',
         'Crime TV Shows, TV Dramas, TV Mysteries',
```

```
In [28]: data['description'].unique()
```

```
Out[28]: array(['Brought together by meaningful meals in the past and present, a doctor and a chef are reacquainted when they begin working at a hospice ward.',
         "From Sierra de las Minas to Esquipulas, explore Guatemala's cultural and geological wealth, including ancient Mayan cities and other natural wonders.",
         'A goofy copywriter unwittingly convinces the Indian cricket team that she’s their lucky mascot, to the dismay of their superstition-shunning captain.',
         ...,
         'When he causes a car accident that leaves a young woman with amnesia, a man finds himself telling incredible lies to make her and her family happy.',
         'In each episode, four celebrities join host Jon Favreau for dinner and share revealing stories about both show business and their personal lives.',
         'While covering a story in New York City, a Seattle-based reporter uncovers a link between two missing women that changes her article entirely.'],
        dtype=object)
```

4 DATA ANALYSIS AND DATA VISUALIZATION

i) Plot the count of different types of shows availble on netflix

In [29]:

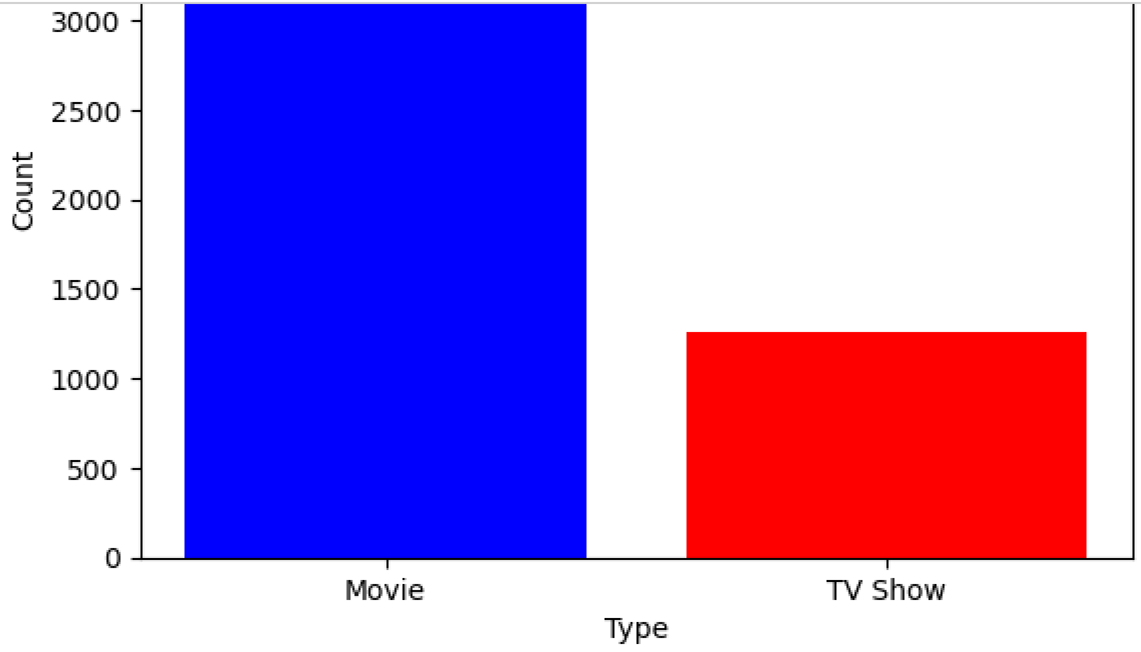
```
type_counts = data['type'].value_counts()
type_counts
```

Out[29]:

```
Movie      3931
TV Show    1255
Name: type, dtype: int64
```

In [63]:

```
# Define colors (matching Netflix pattern)
colors = ['blue', 'red']

# Create bar chart
plt.bar(type_counts.index, type_counts.values, color=colors)

# Add title and labels
plt.title('Count of Different Shows in NETFLIX')
plt.xlabel('Type')
plt.ylabel('Count')

# Show the plot
plt.show()
```



observations from the graph:

1. Netflix contains more movies than TV shows.
2. It indicates that Movies has more audience than TV Shows.

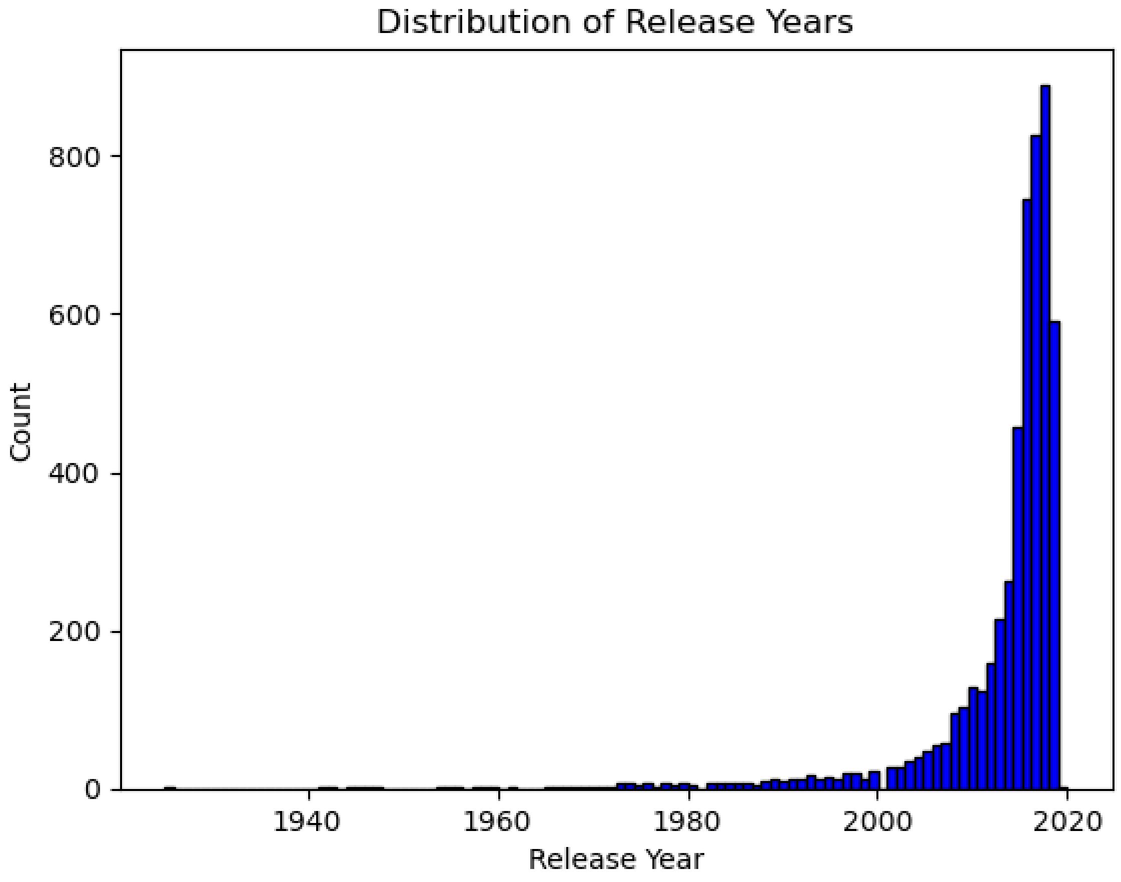
In [30]:
```
data['type'].value_counts()
```

Out[30]:
```
Movie      3931
TV Show    1255
Name: type, dtype: int64
```

ii) Plot the Distribution of Release years

In [31]:
```
plt.hist(data['release_year'], bins=100, color='b', edgecolor='black')

# Add title and labels
plt.title("Distribution of Release Years")
plt.xlabel("Release Year")
plt.ylabel("Count")
```

Out[31]:
```
Text(0, 0.5, 'Count')
```



observations from the above graph:

1. Most number of releases are during the period 2000-2020
2. Highest no: of releases are in the year 2018

In [32]:

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

Out[32]:

```
2018    890
2017    826
2016    746
2019    590
2015    458
       ...
1969      1
1956      1
2020      1
1947      1
1954      1
Name: release_year, Length: 70, dtype: int64
```

iii) Top 10 countries with highest content

In [33]:

```
data['country'].value_counts().head(10)
```

Out[33]:

```
United States     1641
India              693
not spicified      393
United Kingdom     270
Japan              122
South Korea        119
Canada             105
Spain               99
France              78
Mexico              76
Name: country, dtype: int64
```

In [ ]:

```

```

In [34]:

```
data.country.value_counts().head(10)
```

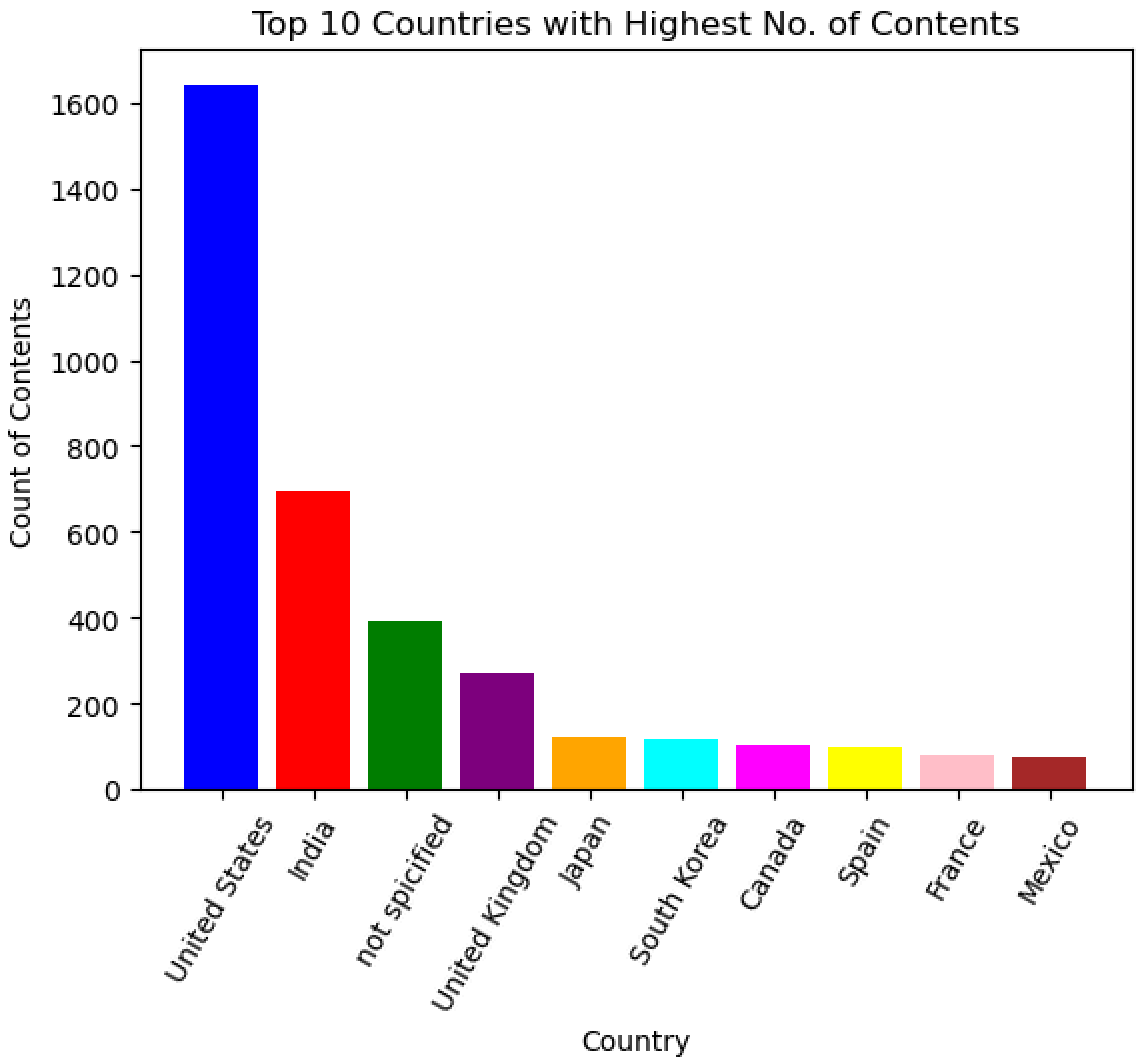
Out[34]:

```
United States     1641
India              693
not spicified      393
United Kingdom     270
Japan              122
South Korea        119
Canada             105
Spain               99
France              78
Mexico              76
Name: country, dtype: int64
```

In [35]:

```
top_countries=data['country'].value_counts().head(10)
```

In [76]:
```
colors = ['blue', 'red', 'green', 'purple', 'orange', 'cyan', 'magenta', 'yellow', 'pink', 'brown']

plt.bar(top_countries.index, top_countries.values, color=colors)

# Rotate x-axis labels for better readability
plt.xticks(rotation=60)

# Add labels and title
plt.ylabel("Count of Contents")
plt.xlabel("Country")
plt.title("Top 10 Countries with Highest No. of Contents")

# Show the plot
plt.show()
```



observations from the graph:

1. Netflix has a greater number of contents from United States.
2. India stands out in the second position.

iV) list out TV shows from India?

In [36]:

```
data[(data['type']=='TV Show')&(data['country']=='India')]['title']
```

Out[36]:

```
361                 Mighty Little Bheem: Diwali
456                             College Romance
460                           Engineering Girls
464                                Girls Hostel
471                                     Inmates
521                               Bard of Blood
826                            Regiment Diaries
847                                   7 (Seven)
898                                  Typewriter
943                                Chhota Bheem
1125                                      Leila
1248                                  Thackeray
1565    Midnight Misadventures With Mallika Dua
1577                                The Calling
1621                                Delhi Crime
1753              Cricket Fever: Mumbai Indians
2166                              Selection Day
2230                        Anjaan: Rural Myths
2301               21 Sarfarosh: Saragarhi 1897
2562              Devlok with Devdutt Pattanaik
2849                            Yeh Meri Family
2868                                      GHOUL
3130                              Dharmakshetra
3131              Raja Rasoi Aur Anya Kahaniyan
3132             Stories by Rabindranath Tagore
3252                  Gabru: Hip Hop Revolution
3329                Anjaan: Special Crimes Unit
3341                   Khan: No. 1 Crime Hunter
3466                                Bh Se Bhade
3493                                    Ramayan
3509      Th Eena Meena Deeka Chase Comedy Show
3556                              Khotey Sikkey
3567                                 Rishta.com
3616                       Darr Sabko Lagta Hai
3642                                   Mahi Way
3726                                     Powder
4674              Fear Files... Har Mod Pe Darr
4678                             Jhansi Ki Rani
4698         The Golden Years with Javed Akhtar
4736                            Classic Legends
4849                               Agent Raghav
4850                             Bhaage Re Mann
4855                          Gangs of Hassepur
4857                           Maharakshak Devi
4886                 Badalte Rishton Ki Dastaan
4900           Khelti Hai Zindagi Aankh Micholi
4904                         Maharakshak: Aryan
4906                               Razia Sultan
Name: title, dtype: object
```

V) list out Movies from India?

In [37]:
```
data[(data['type']=='Movie')&(data['country']=='India')]['title']
```

Out[37]:
```
2             The Zoya Factor
24        Evvarikee Cheppoddu
34                    Pranaam
69                  Dorasaani
78                  Mallesham
               ...           
5407                    Harud
5420                   Umrika
5478                   Dhanak
5485              John & Jane
5545       X: Past Is Present
Name: title, Length: 645, dtype: object
```

Vi) Top 10 Directors with the highest no: of contents

In [38]:
```
data['Directors'].value_counts().head(10)
```

Out[38]:
```
not spicified             1291
Raúl Campos, Jan Suter      18
Marcus Raboy                14
Jay Karas                   13
Jay Chapman                 12
Martin Scorsese              8
Steven Spielberg             8
Johnnie To                   8
Shannon Hartman              7
Hakan Algül                  7
Name: Directors, dtype: int64
```

In [40]:
```
# Remove rows where 'Directors' is 'not spicified' and then count values
top_directors = data[data['Directors'] != 'not spicified']['Directors'].value_counts().head(10)

# Display the result
print(top_directors)
```

Out[40]:
```
Raúl Campos, Jan Suter    18
Marcus Raboy              14
Jay Karas                 13
Jay Chapman               12
Martin Scorsese            8
Steven Spielberg           8
Johnnie To                 8
Hakan Algül                7
Ryan Polito                7
Cathy Garcia-Molina        7
Name: Directors, dtype: int64
```

Vii) List all the movies of Mohanlal (actor)?

In [41]:
```
data[data['cast'].str.contains('Mohanlal')]
```

Out[41]:

	show_id	title	Directors	cast	country	date_added	release_year	rating	duration	listed_in	description	type
2469	70241088	Grandmaster	B. Unnikrishnan	Mithra Kurian, Anoop Menon, Mohanlal, Narein, ...	India	November 1, 2018	2012	TV-14	133 min	International Movies, Thrillers	Anticipating the next move in a serial killer'...	Movie
3615	60036613	Company	Ram Gopal Varma	Ajay Devgn, Vivek Oberoi, Mohanlal, Manisha Ko...	India	March 1, 2018	2002	TV-MA	137 min	Action & Adventure, Dramas, International Movies	Following a misunderstanding, a gangster and h...	Movie
4043	70236586	Tezz	Priyadarshan	Anil Kapoor, Ajay Devgn, Boman Irani, Kangna R...	United Kingdom, India	December 1, 2017	2012	TV-MA	121 min	Action & Adventure, International Movies, Musi...	This high-octane thriller tells the story of a...	Movie

Viii) How many Movies got the 'TV-14' rating in India?

In [42]:
```
data[(data['rating']=='TV-14') & (data['country']=='India')]
```

Out[42]:

	show_id	title	Directors	cast	country	date_added	release_year	rating	duration	listed_in	description	type
2	81213894	The Zoya Factor	Abhishek Sharma	Sonam Kapoor, Dulquer Salmaan, Sanjay Kapoor, ...	India	November 30, 2019	2019	TV-14	135 min	Comedies, Dramas, International Movies	A goofy copywriter unwittingly convinces the I...	Movie
24	81194544	Evvarikee Cheppoddu	Basava Shankar Eeday	Rakesh Varre, Gargeyi, Vamsi raj Nekkanti, D P...	India	November 27, 2019	2019	TV-14	134 min	Comedies, International Movies, Romantic Movies	When caste differences throw a wrench into the...	Movie
69	81217738	Dorasaani	KVR Mahendra	Anand Deverakonda, Shivatmika Rajasekhar, Kann...	India	November 20, 2019	2019	TV-14	134 min	Dramas, International Movies, Romantic Movies	A village landlord's daughter and an educated ...	Movie
111	81076114	House Arrest	Shashanka Ghosh, Samit Basu	Ali Fazal, Shriya Pilgaonkar, Jim Sarbh, Barkh...	India	November 15, 2019	2019	TV-14	105 min	Comedies, Independent Movies, International Mo...	A world-weary man's self-imposed home confinem...	Movie
151	80127886	U Turn	Pawan Kumar	Roger Narayan, Shraddha Srinath, Dileep Raj, K...	India	November 7, 2019	2016	TV-14	121 min	International Movies, Thrillers	A reporter must hunt for the truth behind a st...	Movie
...	...	...	...	...	...	...	...	...	...	...	...	...
5258	80139507	Famous in Ahmedabad	Hardik Mehta	not spicified	India	December 15, 2016	2015	TV-14	29 min	Documentaries, International Movies	Follow an 11-year-old boy as he goes from ordi...	Movie
5340	70260514	Mumbai Cha Raja	Manjeet Singh	Rahul Bairagi, Arbaaz Khan, Tejas Parvatkar, D...	India	November 15, 2016	2012	TV-14	77 min	Dramas, Independent Movies, International Movies	This coming-of-age tale follows Rahul, a young...	Movie
5407	70153553	Harud	Aamir Bashir	Mohammad Amir Naji, Shahnawaz Bhat, Shamim Bas...	India	October 1, 2016	2010	TV-14	100 min	Dramas, International Movies	A young Kashmiri man sets out to escape to Pak...	Movie
5420	80038288	Umrika	Prashant Nair	Suraj Sharma, Tony Revolori, Smita Tambe, Adil...	India	October 1, 2016	2015	TV-14	96 min	Dramas, Independent Movies, International Movies	An ambitious man leaves his small Indian villa...	Movie
5485	80115844	John & Jane	Ashim Ahluwalia	not spicified	India	August 15, 2016	2005	TV-14	79 min	Documentaries, International Movies	Truth and fiction blend in this quasi-document...	Movie

369 rows × 12 columns

ix) Which individual country has the highest no: of TV Shows?

In [43]:

```
data[data['type']=='TV Show']['country'].value_counts().head(1)
```

Out[43]:

```
United States    269
Name: country, dtype: int64
```

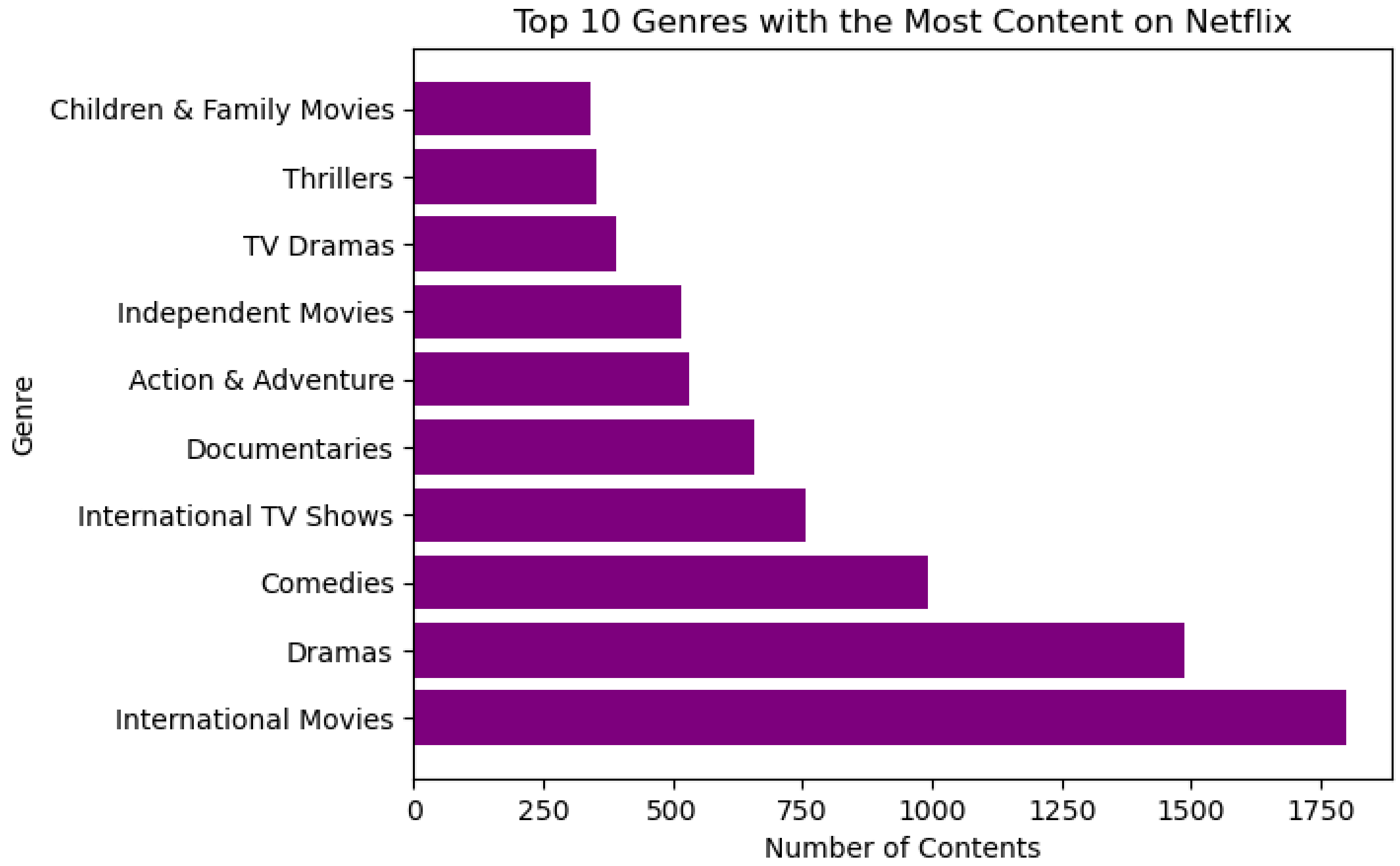
x) Top 10 Genres with the Most Content on Netflix

In [44]:

```
#Extracting unique genres from the 'listed_in' column
all_genres = ', '.join(data['listed_in']).split(', ')

#Get the unique genres by converting the list into a set (removes duplicates) and back into a list
unique_genres = list(set(all_genres))

#Count how many times each genre appears in the list
genre_counts = {}  # Start with an empty dictionary
for genre in unique_genres: #it will iterate the genres one by one(Action,Comedy,Action,Drama,Action,Comedy)
    genre_counts[genre] = all_genres.count(genre)  # it will Count each genre's occurrences in all_gnres list

#Get the top 10 genres with the most content
top_genres = dict(sorted(genre_counts.items(), key=lambda x: x[1], reverse=True)[:10])

#genre_counts.items() => This takes the genre_counts dictionary and converts it into a list of key-value pairs (tuples). example:[('Action', 3), ('Comedy', 2), ('Drama', 1)]
#reverse=True => it would sort in ascending order (lowest count to highest count).
```

In [45]:

```
#Create a horizontal bar plot using Matplotlib
plt.barh(list(top_genres.keys()), list(top_genres.values()), color='purple')

#Add labels and title
plt.xlabel('Number of Contents')
plt.ylabel('Genre')
plt.title('Top 10 Genres with the Most Content on Netflix')

# Show the plot
plt.show()
```



In [ ]:

```
```