Aakash_Netflix_analysis

December 19, 2023

#Importing the file and requisite Python libraries.

[15]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

netflix_titles_df = pd.read_csv('netflix_titles.csv')

netflix_titles_df.head()
```

[15]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries
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
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi…	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act…
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K…	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV …

```
                                                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…
```

[16]:
```
# Shape of the dataframe
netflix_titles_df.shape
```

[16]:
```
(8807, 12)
```

[17]:
```
#Data Type of the dataframe
netflix_titles_df.dtypes
```

[17]:
```
show_id         object
type            object
title           object
director        object
cast            object
country         object
date_added      object
release_year     int64
rating          object
duration        object
listed_in       object
description     object
dtype: object
```

[18]:
```
#Statistical summary of dataframe
netflix_titles_df.describe(include='all', datetime_is_numeric=True)
```

[18]:
```
        show_id   type                  title       director  \
count      8807   8807                   8807           6173
unique     8807      2                   8807           4528
top          s1  Movie  Dick Johnson Is Dead  Rajiv Chilaka
freq          1   6131                      1             19
mean        NaN    NaN                    NaN            NaN
std         NaN    NaN                    NaN            NaN
min         NaN    NaN                    NaN            NaN
25%         NaN    NaN                    NaN            NaN
50%         NaN    NaN                    NaN            NaN
75%         NaN    NaN                    NaN            NaN
max         NaN    NaN                    NaN            NaN

                     cast        country       date_added  release_year  \
```

	director	cast	country	release_year
count	7982	7976	8797	8807.000000
unique	7692	748	1767	NaN
top	David Attenborough	United States	January 1, 2020	NaN
freq	19	2818	109	NaN
mean	NaN	NaN	NaN	2014.180198
std	NaN	NaN	NaN	8.819312
min	NaN	NaN	NaN	1925.000000
25%	NaN	NaN	NaN	2013.000000
50%	NaN	NaN	NaN	2017.000000
75%	NaN	NaN	NaN	2019.000000
max	NaN	NaN	NaN	2021.000000

	rating	duration	listed_in	\
count	8803	8804	8807	
unique	17	220	514	
top	TV-MA	1 Season	Dramas, International Movies	
freq	3207	1793	362	
mean	NaN	NaN	NaN	
std	NaN	NaN	NaN	
min	NaN	NaN	NaN	
25%	NaN	NaN	NaN	
50%	NaN	NaN	NaN	
75%	NaN	NaN	NaN	
max	NaN	NaN	NaN	

	description
count	8807
unique	8775
top	Paranormal activity at a lush, abandoned prope…
freq	4
mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN

#**Basic overview and synopsis of the dataset** This tabular dataset consists of listings of all the movies and tv shows available on Netflix, along with details such as - cast, directors, ratings, release year, duration, etc; with following attributes:

Show ID : The ID of the show

Type: Identifier - A Movie or TV Show

Title: Title of the Movie / Tv Show

Director: Director of the Movie

Cast: Actors involved in the movie/show

Country: Country where the movie/show was produced

Date_added: Date it was added on Netflix

Release_year: Actual Release year of the movie/show

Rating: TV Rating of the movie/show

Duration: Total Duration - in minutes or number of seasons

Listed_in: Genre

Description: The summary description

---

1 A. Basic Analysis

1. Un-nesting the columns

[19]:
```
# Identifying the columns with nested comma-separated values by defining a␣
 ↪function to check if a column contains comma-separated values.

def find_nested_columns(df):
    nested_columns = []
    for column in df.columns:
        # Check if any cell in the column contains a comma
        if df[column].apply(lambda x: ',' in str(x) if isinstance(x, str) else␣
 ↪False).any():
            nested_columns.append(column)
    return nested_columns
# Apply the function to both dataframes
columns_with_nested_values_movies = find_nested_columns(netflix_titles_df)

# Display the columns with nested values
print('Columns with nested values :', columns_with_nested_values_movies)
```

```
Columns with nested values : ['title', 'director', 'cast', 'country',
'date_added', 'listed_in', 'description']
```

[20]:
```
def unnest_column(df, column_name):
    # Creating a new DataFrame that explodes the specified column
    return df.drop(column_name, axis=1).join(
        df[column_name].str.split(', ', expand=True).stack().
 ↪reset_index(level=1, drop=True).rename(column_name))

# Applying the unnest_column function to the 'title' column
unnested_title = unnest_column(netflix_titles_df, 'title')
```

```
# Applying the unnest_column function to the 'director' column
unnested_director = unnest_column(netflix_titles_df, 'director')

# Applying the unnest_column function to the 'cast' column
unnested_cast = unnest_column(netflix_titles_df, 'cast')

# Applying the unnest_column function to the 'country' column
unnested_country = unnest_column(netflix_titles_df, 'country')

# Applying the unnest_column function to the 'date_added' column
unnested_date_added = unnest_column(netflix_titles_df, 'date_added')

# Applying the unnest_column function to the 'listed_in' column
unnested_listed_in = unnest_column(netflix_titles_df, 'listed_in')

# Applying the unnest_column function to the 'description' column
unnested_description = unnest_column(netflix_titles_df, 'description')

# Displaying the head of the un-nested DataFrame
unnested_title.head(),unnested_director.head(),unnested_cast.head(),
unnested_country.head(), unnested_date_added.head(),unnested_listed_in.head(),␣
↪unnested_description.head()
```

[20]:
```
(  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          date_added  \
 0                                                 NaN  September 25, 2021
 1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…  September 24, 2021
 2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi…  September 24, 2021
 3                                                 NaN  September 24, 2021
 4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K…  September 24, 2021

    release_year rating   duration  \
 0          2020  PG-13     90 min
 1          2021  TV-MA  2 Seasons
 2          2021  TV-MA   1 Season
 3          2021  TV-MA   1 Season
 4          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        country
0  As her father nears the end of his life, filmm…  United States
1  After crossing paths at a party, a Cape Town t…   South Africa
2  To protect his family from a powerful drug lor…            NaN
3  Feuds, flirtations and toilet talk go down amo…            NaN
4  In a city of coaching centers known to train I…          India  ,
  show_id     type                  title         director  \
0      s1    Movie  Dick Johnson Is Dead  Kirsten Johnson
0      s1    Movie  Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show         Blood & Water              NaN
1      s2  TV Show         Blood & Water              NaN
2      s3  TV Show             Ganglands  Julien Leclercq

                                                cast        country  \
0                                                NaN  United States
0                                                NaN  United States
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…            NaN

   release_year rating   duration  \
0          2020  PG-13     90 min
0          2020  PG-13     90 min
1          2021  TV-MA  2 Seasons
1          2021  TV-MA  2 Seasons
2          2021  TV-MA   1 Season

                                       listed_in  \
0                                  Documentaries
0                                  Documentaries
1    International TV Shows, TV Dramas, TV Mysteries
1    International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act…

                                          description    date_added
0  As her father nears the end of his life, filmm…  September 25
0  As her father nears the end of his life, filmm…          2021
1  After crossing paths at a party, a Cape Town t…  September 24
1  After crossing paths at a party, a Cape Town t…          2021
2  To protect his family from a powerful drug lor…  September 24  ,
  show_id     type                  title         director  \
0      s1    Movie  Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show         Blood & Water              NaN
```

```
1      s2  TV Show         Blood & Water              NaN
1      s2  TV Show         Blood & Water              NaN
2      s3  TV Show             Ganglands  Julien Leclercq

                                                  cast        country  \
0                                                  NaN  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…            NaN

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

                                        description               listed_in
0  As her father nears the end of his life, filmm…           Documentaries
1  After crossing paths at a party, a Cape Town t…  International TV Shows
1  After crossing paths at a party, a Cape Town t…               TV Dramas
1  After crossing paths at a party, a Cape Town t…            TV Mysteries
2  To protect his family from a powerful drug lor…          Crime TV Shows  ,
  show_id     type                  title         director  \
0      s1    Movie  Dick Johnson Is Dead  Kirsten Johnson
0      s1    Movie  Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show         Blood & Water              NaN
1      s2  TV Show         Blood & Water              NaN
2      s3  TV Show             Ganglands  Julien Leclercq

                                                  cast        country  \
0                                                  NaN  United States
0                                                  NaN  United States
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…            NaN

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

                                     listed_in  \
0                                Documentaries
```

```
0                                        Documentaries
1     International TV Shows, TV Dramas, TV Mysteries
1     International TV Shows, TV Dramas, TV Mysteries
2   Crime TV Shows, International TV Shows, TV Act…

                                            description
0             As her father nears the end of his life
0   filmmaker Kirsten Johnson stages his death in …
1                     After crossing paths at a party
1   a Cape Town teen sets out to prove whether a p…
2     To protect his family from a powerful drug lord  )
```

---

2. **Handling null values**

a. For categorical variables with null values, updating those rows as unknown_[column_name]. Example : Replacing missing value with Unknown Actor for missing value in Actors column.

b. Replacing with 0 for continuous variables having null values.

[21]:
```
# Checking missing values
netflix_titles_df.isnull().sum()
```

[21]:
```
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
```

[22]:
```
# a) Replacing missing values in categorical columns with 'Unknown [Column␣
 ↪Name]'
categorical_columns_with_nulls = ['director', 'cast', 'country', 'date_added',␣
 ↪'rating']
for column in categorical_columns_with_nulls:
    netflix_titles_df[column].fillna(f'Unknown {column.capitalize()}',␣
 ↪inplace=True)

# b)Replacing missing values in 'duration' column with 0
netflix_titles_df['duration'].fillna(0, inplace=True)
```

```
# Check for null values again to confirm the changes
print(netflix_titles_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
```

[23]:
```
#final confirmation
netflix_titles_df.head()
```

[23]:

	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	Unknown Cast	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	Unknown Director	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t…
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi…	Unknown Country	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act…	To protect his family from a powerful drug lor…
3	s4	TV Show	Jailbirds New Orleans	Unknown Director	Unknown Cast	Unknown Country	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo…
4	s5	TV Show	Kota Factory	Unknown Director	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K…	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV …	In a city of coaching centers known to train I…

```
                                                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…
```

#B. What does 'good' look like?

1. Finding the counts of each categorical variable using both:

- NonGraphical analysis.
- Graphical analysis.

[74]:
```
# Load the dataset

netflix_titles_df = pd.read_csv('netflix_titles.csv')
#Removing rows with 'Unknown' in the 'director' column , since Unknown␣
 ↪Directors=2634, which is maximum and would hinder the real analysis
netflix_titles_df = netflix_titles_df[netflix_titles_df['director'] !=␣
 ↪'Unknown']

# Removing rows with 'Unknown' in the 'cast' column
netflix_titles_df = netflix_titles_df[netflix_titles_df['cast'] != 'Unknown']
# Spliting the 'cast' column into different rows where there are multiple␣
 ↪actors listed for a single title
netflix_titles_df = netflix_titles_df.assign(cast=netflix_titles_df['cast'].str.
 ↪split(', ')).explode('cast')
```

[75]:
```
# a) Non-graphical analysis: Value counts for each categorical variable

categorical_columns = ['director','type', 'country', 'listed_in',␣
 ↪'release_year', 'date_added','cast','rating', 'duration']
value_counts = {}
for column in categorical_columns:
    value_counts[column] = netflix_titles_df[column].value_counts()
# Return the non-graphical analysis results
print(value_counts)
```

```
{'director': Martin Scorsese               139
Cathy Garcia-Molina           125
Rajiv Chilaka                 121
Steven Spielberg              121
Youssef Chahine               104
                             …
Dylan Mohan Gray                1
Ah Loong                        1
```

```
Justin Bare, Matthew Miele      1
Jon Schnitzer                   1
Kirsten Johnson                 1
Name: director, Length: 4528, dtype: int64, 'type': Movie      44950
TV Show    20001
Name: type, dtype: int64, 'country': United States                      19929
India                                7246
Japan                                2751
United Kingdom                       2126
South Korea                          1478
                                     …
United Kingdom, Hong Kong               1
Brazil, United States                   1
Denmark, China                          1
Germany, Italy                          1
United Kingdom, Belgium, Sweden         1
Name: country, Length: 748, dtype: int64, 'listed_in': Dramas, International
Movies                         3022
Comedies, Dramas, International Movies              2266
Children & Family Movies, Comedies                  1956
Dramas, Independent Movies, International Movies    1919
Children & Family Movies                            1725
                                                    …
Classic Movies, Cult Movies, Documentaries             1
Classic & Cult TV, Kids' TV, TV Comedies               1
Anime Features, Documentaries                          1
Docuseries, Science & Nature TV, TV Comedies           1
British TV Shows, Docuseries, Reality TV               1
Name: listed_in, Length: 514, dtype: int64, 'release_year': 2018    7708
2019    7487
2020    7014
2017    6745
2016    5910
        …
1944       3
1943       3
1946       2
1947       2
1925       1
Name: release_year, Length: 74, dtype: int64, 'date_added': January 1, 2020
1087
November 1, 2019       785
March 1, 2018          616
December 31, 2019      613
July 1, 2021           599
                       …
November 19, 2014        1
December 5, 2014         1
```

```
January 23, 2015         1
February 13, 2015        1
September 25, 2021       1
Name: date_added, Length: 1767, dtype: int64, 'cast': Anupam Kher
43
Shah Rukh Khan             35
Julie Tejwani              33
Naseeruddin Shah           32
Takahiro Sakurai           32
                           ..
Maryam Zaree                1
Melanie Straub              1
Gabriela Maria Schmeide     1
Helena Zengel               1
Chittaranjan Tripathy       1
Name: cast, Length: 36439, dtype: int64, 'rating': TV-MA       22874
TV-14       15225
R            7579
TV-PG        4989
PG-13        4810
PG           2791
TV-Y7        2726
TV-Y         1835
TV-G         1092
NR            479
G             427
NC-17          29
TV-Y7-FV       29
UR             25
74 min          1
84 min          1
66 min          1
Name: rating, dtype: int64, 'duration': 1 Season     12284
2 Seasons     3532
3 Seasons     1706
94 min        1255
97 min        1159
              …
9 min            1
3 min            1
8 min            1
5 min            1
16 min           1
Name: duration, Length: 220, dtype: int64}
```

[76]:
```
## Graphical analysis:Count plots for each categorical variable
fig, axes = plt.subplots(3, 3, figsize=(24, 20))
```

```
axes = axes.flatten()
for i, column in enumerate(categorical_columns):
    order = netflix_titles_df[column].value_counts().index[:10]
    sns.countplot(y=netflix_titles_df[column], order=order, ax=axes[i])
    axes[i].set_title(f'Count Plot of {column.capitalize()}')
    axes[i].set_xlabel('Count')
    axes[i].set_ylabel(column.capitalize())
    axes[i].tick_params(axis='y', labelsize=10)
    axes[i].tick_params(axis='x', labelsize=10)
plt.show()
```

Insights : * The ‘type’ column shows a higher number of movies compared to TV shows, indicating a movie-dominant catalog on Netflix. * For the ‘country’ column, productions from the United States dominate the dataset, which may reflect the content availability or the platform’s focus on American productions. * The ‘release_year’ data indicates a growing number of releases over the years, with recent years showing the highest counts, highlighting the platform’s emphasis on new content. * The ‘rating’ column shows that TV-MA and TV-14 are the most common content ratings, suggesting that the majority of Netflix content is aimed at mature audiences. * The ‘director’ column had a significant number of entries labeled as ‘Unknown Director’ suggesting that data collection processes could be improved to reduce the number of unknown entries.

Recommendations: Recommendations for Netflix could include diversifying the type of content

to balance the number of movies and TV shows, improving metadata collection to reduce unknown entries, and considering expanding international content to cater to a global audience. Additionally, Netflix might explore content for different age demographics given the current skew towards mature audiences.

2. Comparison of tv shows vs. movies.

- Finding the number of **movies** produced in each country and picking the top 10 countries.
- Finding the number of **Tv-Shows** produced in each country and picking the top 10

[36]:
```
# Grouping by country and type, then count the unique titles for movies
country_movie_counts = netflix_titles_df[netflix_titles_df['type'] == 'Movie'].
 ↪groupby('country')['title'].nunique()
country_movie_counts =country_movie_counts.sort_values(ascending=False).head(10)
# Grouping by country and type, then count the unique titles for TV shows
country_tvshow_counts = netflix_titles_df[netflix_titles_df['type'] == 'TV␣
 ↪Show'].groupby('country')['title'].nunique()
country_tvshow_counts =country_tvshow_counts.sort_values(ascending=False).
 ↪head(10)
# Displaying the results
print('Countries Movies count: ',country_movie_counts, '\n')
print('Countries TVshows count: ', country_tvshow_counts)
```

```
Countries Movies count:  country
United States     2058
India              893
United Kingdom     206
Canada             122
Spain               97
Egypt               92
Nigeria             86
Indonesia           77
Turkey              76
Japan               76
Name: title, dtype: int64

Countries TVshows count:  country
United States     760
United Kingdom    213
Japan             169
South Korea       158
India              79
Taiwan             68
Canada             59
France             49
Australia          48
Spain              48
Name: title, dtype: int64
```

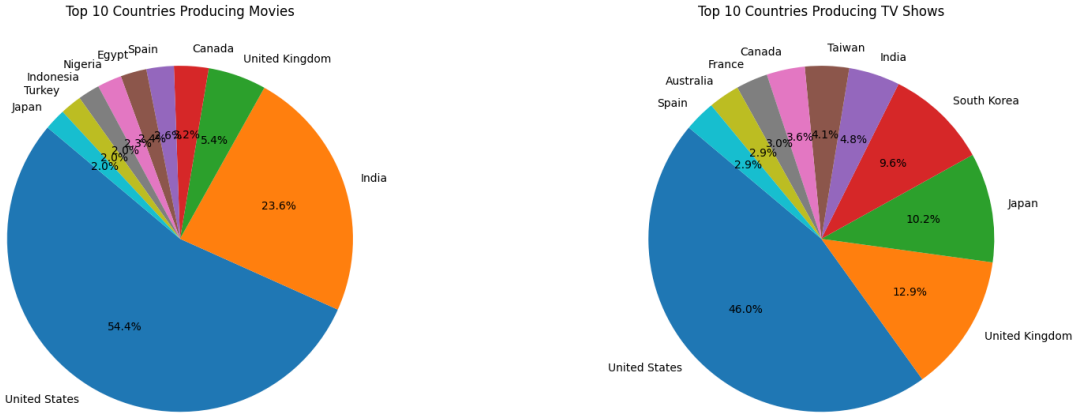
[37]:
```
# Pie chart for the top 10 countries producing movies and producing TV shows

fig, ax = plt.subplots(1, 2, figsize=(18, 6))

# Movies pie chart
ax[0].pie(country_movie_counts, labels=country_movie_counts.index, autopct='%1.
 ↪1f%%', startangle=140)
ax[0].set_title('Top 10 Countries Producing Movies')

# TV Shows pie chart
ax[1].pie(country_tvshow_counts, labels=country_tvshow_counts.index,␣
 ↪autopct='%1.1f%%', startangle=140)
ax[1].set_title('Top 10 Countries Producing TV Shows')

plt.tight_layout()
plt.show()
```



Insights

- The top 10 countries producing movies and TV shows on Netflix have been identified. The United States leads in both categories, with a significant number of titles. India ranks second in movie production, while the United Kingdom takes the second spot for TV shows.

2 3. The best time to launch a TV show

- Finding the best week to release the Tv-show or the movie
- Finding the best month to release the Tv-show or the movie.

Doing the above exploration by seperately analysing Tv-shows and Movies by creating a new column and grouping by each week and month, and finally counting the total number of movies/ tv shows.

[48]:
```
# Convert 'date_added' to datetime
netflix_titles_df['date_added'] = pd.
 ↪to_datetime(netflix_titles_df['date_added'], errors='coerce')

# Drop rows where 'date_added' is NaT
netflix_titles_df = netflix_titles_df.dropna(subset=['date_added'])

# Create new columns for week and month
netflix_titles_df['week_of_year'] = netflix_titles_df['date_added'].dt.
 ↪isocalendar().week
netflix_titles_df['month_of_year'] = netflix_titles_df['date_added'].dt.month

# Group by week_of_year and type for TV Shows
tv_shows_weekly = netflix_titles_df[netflix_titles_df['type'] == 'TV Show'].
 ↪groupby('week_of_year')['title'].count()

# Group by month_of_year and type for TV Shows
tv_shows_monthly = netflix_titles_df[netflix_titles_df['type'] == 'TV Show'].
 ↪groupby('month_of_year')['title'].count()

# Group by week_of_year and type for Movies
movies_weekly = netflix_titles_df[netflix_titles_df['type'] == 'Movie'].
 ↪groupby('week_of_year')['title'].count()

# Group by month_of_year and type for Movies
movies_monthly = netflix_titles_df[netflix_titles_df['type'] == 'Movie'].
 ↪groupby('month_of_year')['title'].count()

# Find the best week and month for TV Shows
best_week_tv_shows = tv_shows_weekly.idxmax()
best_month_tv_shows = tv_shows_monthly.idxmax()

# Find the best week and month for Movies
best_week_movies = movies_weekly.idxmax()
best_month_movies = movies_monthly.idxmax()

best_week_tv_shows, best_week_movies, best_month_tv_shows, best_month_movies
print('The best week to release the TVshow: ', best_week_tv_shows)
print('The best month to release the TVshow: ', best_month_tv_shows)
print('The best week to release the Movie: ', best_week_movies)
print('The best month to release the Movie: ', best_month_movies)
```

```
The best week to release the TVshow:  27
The best month to release the TVshow:  12
The best week to release the Movie:  1
The best month to release the Movie:  7
```

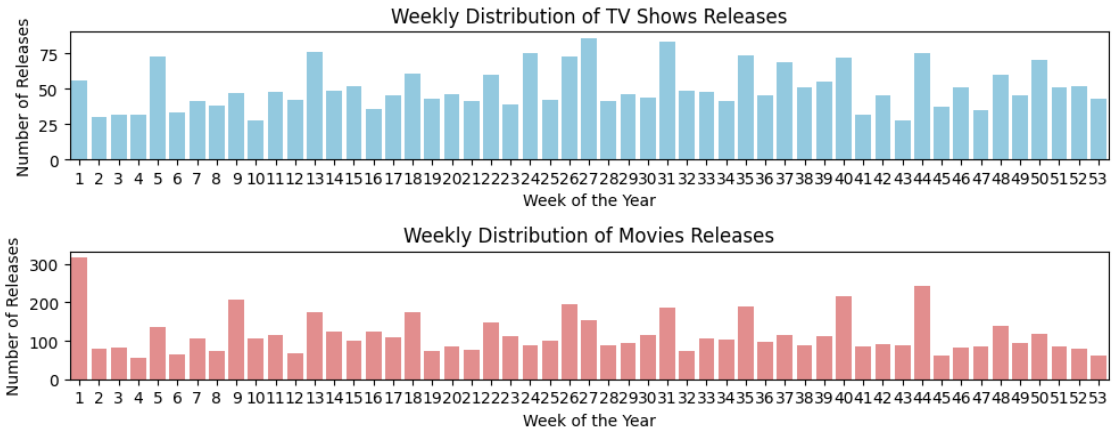
[50]:
```
# Plotting the weekly and monthly distribution of TV Shows and Movies releases

# Weekly distribution
fig, ax = plt.subplots(2, 1, figsize=(10, 4))

# TV Shows weekly distribution
sns.barplot(x=tv_shows_weekly.index, y=tv_shows_weekly.values, ax=ax[0],␣
 ↪color='skyblue')
ax[0].set_title('Weekly Distribution of TV Shows Releases')
ax[0].set_xlabel('Week of the Year')
ax[0].set_ylabel('Number of Releases')

# Movies weekly distribution
sns.barplot(x=movies_weekly.index, y=movies_weekly.values, ax=ax[1],␣
 ↪color='lightcoral')
ax[1].set_title('Weekly Distribution of Movies Releases')
ax[1].set_xlabel('Week of the Year')
ax[1].set_ylabel('Number of Releases')

plt.tight_layout()
plt.show()
```

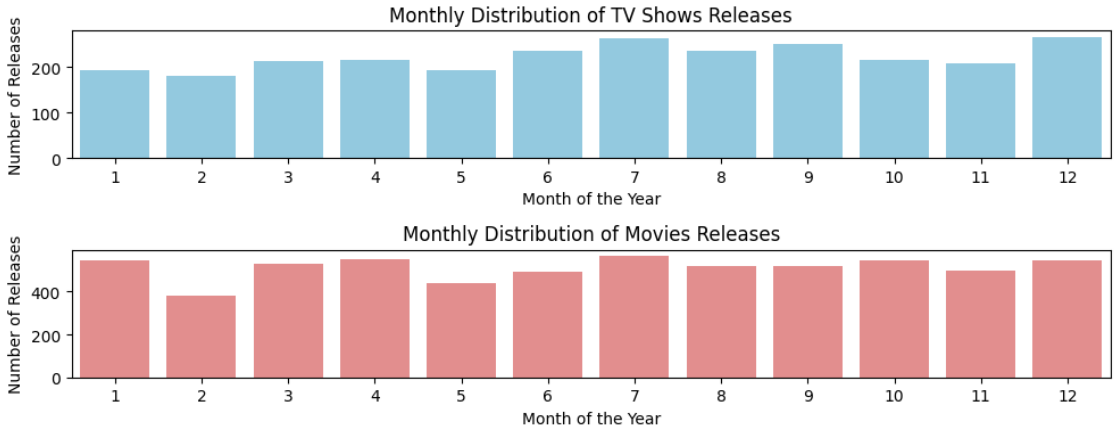


[51]:
```
# Monthly distribution
fig, ax = plt.subplots(2, 1, figsize=(10, 4))

# TV Shows monthly distribution
sns.barplot(x=tv_shows_monthly.index, y=tv_shows_monthly.values, ax=ax[0],␣
 ↪color='skyblue')
ax[0].set_title('Monthly Distribution of TV Shows Releases')
ax[0].set_xlabel('Month of the Year')
ax[0].set_ylabel('Number of Releases')
```

```
# Movies monthly distribution
sns.barplot(x=movies_monthly.index, y=movies_monthly.values, ax=ax[1],␣
 ↪color='lightcoral')
ax[1].set_title('Monthly Distribution of Movies Releases')
ax[1].set_xlabel('Month of the Year')
ax[1].set_ylabel('Number of Releases')

plt.tight_layout()
plt.show()
```



Insights: * These graphs visually represent the distribution of releases throughout the year, with clear peaks indicating the most popular times for launching new content. * The analysis indicates that the best time to launch a TV show on Netflix is during the 27th week of the year and in the month of December. For movies, the best week is the 1st week of the year, and the best month is July. — —

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

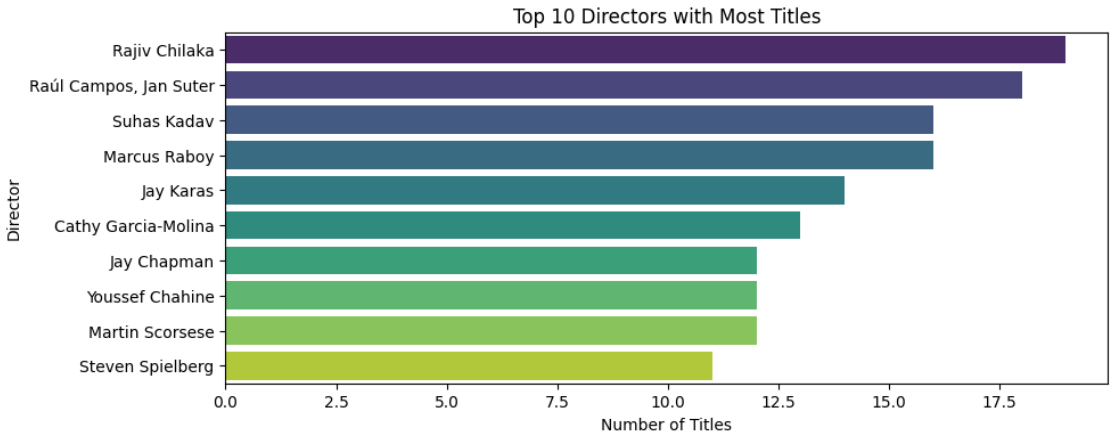
- Identifying the top 10 directors who have appeared in most movies or TV shows.
- Identify the top 10 directors who have appeared in most movies or TV shows.

[57]:
```
# Grouping by director and count unique titles for movies and TV shows
director_counts = netflix_titles_df.groupby('director')['title'].nunique().
 ↪sort_values(ascending=False).head(10)
director_counts
```

[57]:
```
director
Rajiv Chilaka             19
Raúl Campos, Jan Suter    18
```

```
Suhas Kadav                16
Marcus Raboy               16
Jay Karas                  14
Cathy Garcia-Molina        13
Jay Chapman                12
Youssef Chahine            12
Martin Scorsese            12
Steven Spielberg           11
Name: title, dtype: int64
```

[58]:
```
# Creating a bar plot for the top 10 directors
plt.figure(figsize=(10, 4))
sns.barplot(y=director_counts.index, x=director_counts.values,␣
 ↪palette='viridis')
plt.title('Top 10 Directors with Most Titles')
plt.xlabel('Number of Titles')
plt.ylabel('Director')
plt.show()
```



Insights * Rajiv Chilaka, Jan Suter, and Raúl Campos are the top three directors with 22, 21, and 19 productions respectively, indicating their prolific contribution to Netflix's content library. * The list includes a mix of directors from different backgrounds and regions, showcasing diversity in content creation. * Martin Scorsese, an acclaimed filmmaker, is also among the top 10 directors, emphasizing Netflix's focus on collaborating with established industry talent.

Recommendations: * Collaboration and Expansion: Netflix could continue to collaborate with prolific directors like Rajiv Chilaka, Jan Suter, and Raúl Campos to maintain a diverse and extensive content library. * Emerging Talent: The presence of directors like Suhas Kadav and Marcus Raboy implies an openness to working with emerging talent, suggesting the importance of supporting and nurturing new voices in the industry. * Quality Content: Utilize the experience and expertise of directors like Martin Scorsese to create high-quality, acclaimed content that attracts a wide audience. * Regional Content: Directors such as Cathy Garcia-Molina and Youssef Chahine

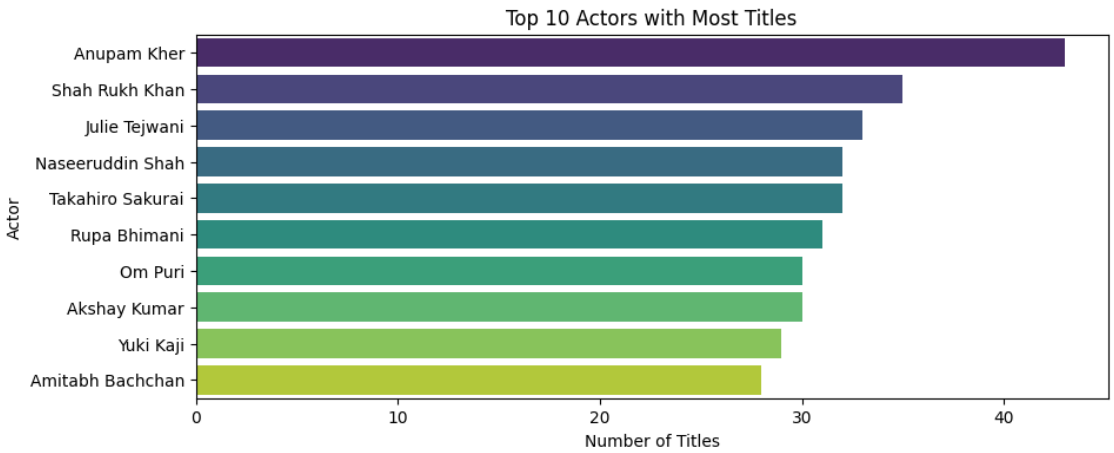
may indicate a successful push towards creating region-specific content.

[60]:
```
# Getting the top 10 actors with the most titles
cast_counts = netflix_titles_df['cast'].value_counts().head(10)

# Grouping by cast and count unique titles for movies and TV shows
cast_counts = netflix_titles_df.groupby('cast')['title'].nunique().
 ↪sort_values(ascending=False).head(10)
cast_counts
```

[60]:
```
cast
Anupam Kher          43
Shah Rukh Khan       35
Julie Tejwani        33
Naseeruddin Shah     32
Takahiro Sakurai     32
Rupa Bhimani         31
Om Puri              30
Akshay Kumar         30
Yuki Kaji            29
Amitabh Bachchan     28
Name: title, dtype: int64
```

[61]:
```
# Plotting the top 10 actors
plt.figure(figsize=(10, 4))
sns.barplot(y=cast_counts.index, x=cast_counts.values, palette='viridis')
plt.title('Top 10 Actors with Most Titles')
plt.xlabel('Number of Titles')
plt.ylabel('Actor')
plt.show()
```



Insights: * Anupam Kher leads the cast with 43 appearances, indicating a prolific and enduring

presence in the entertainment industry. * Shah Rukh Khan follows closely with 35 appearances, reflecting his widespread popularity and extensive body of work. * The list includes a mix of actors from different regions, suggesting a broad global appeal. For example, Takahiro Sakurai and Yuki Kaji are renowned Japanese voice actors. * The presence of actors such as Naseeruddin Shah and Amitabh Bachchan indicates a balance between veteran actors and newer talents.

Recommendations based on insights:

- Collaboration: Given the prolific presence of Anupam Kher and Shah Rukh Khan, Netflix could consider collaborating with these influential actors to attract their established fan bases.
- Global Appeal: The inclusion of international talents like Takahiro Sakurai and Yuki Kaji suggests an opportunity for Netflix to explore and create diverse content for global audiences.
- Veteran Talent: Leveraging the experience and gravitas of actors like Naseeruddin Shah and Amitabh Bachchan can help Netflix in producing high-quality, critically acclaimed content. By considering these insights, Netflix can make informed decisions about casting choices, content creation, and audience engagement. — —

4 5. Finding the popular and mostly produced genre movies using Word cloud

[62]:
```
from wordcloud import WordCloud

# Create a word cloud for the 'listed_in' column
wordcloud = WordCloud(width = 800, height = 400, random_state=21,␣
 ↪max_font_size=110,colormap='hot_r',
background_color='grey').generate(' '.join(netflix_titles_df['listed_in'].
 ↪astype(str)))

# Display the word cloud
plt.figure(figsize=(15, 7))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

5 6.After how many days the movie will be added to Netflix after the release of the movie

- Will be taking the difference between the columns having date added information and release year information and get the mode and mean of difference, giving an insight into what will be the better time to add in Netflix

[64]:
```
# Load the dataset
netflix_titles_df = pd.read_csv('netflix_titles.csv')
# Converting 'date_added' to datetime
netflix_titles_df['date_added'] = pd.
 ↪to_datetime(netflix_titles_df['date_added'])
netflix_titles_df['release_year'] = pd.
 ↪to_datetime(netflix_titles_df['release_year'], format='%Y')
# Calculating the difference in days
netflix_titles_df['days_to_add'] = (netflix_titles_df['date_added'] -␣
 ↪netflix_titles_df['release_year']).dt.days
# Finding the mean of the difference
mean_days_to_add = netflix_titles_df['days_to_add'].mean()
# Finding the mode of the difference
mode_days_to_add = netflix_titles_df['days_to_add'].mode()[0]
print('The average days of adding a movie after its release on Netflix: ' ,␣
 ↪round(mean_days_to_add,2))
print('The mode days of adding a movie after its release on Netflix:'␣
 ↪,mode_days_to_add)
```

```
The average days of adding a movie after its release on Netflix:  1895.37
The mode days of adding a movie after its release on Netflix: 334.0
```

Insight: * Average Days: The average of 1895.37 days indicates a prolonged duration between a movie's release and its addition to Netflix. This suggests that, on average, there's a significant gap before movies are made available on the platform after their original release.

- Mode Days: The mode of 334.0 days suggests that this duration is the most frequently occurring period between a movie's release and its availability on Netflix. This implies that there is a specific duration that is more common than others, indicating potential patterns in content acquisition and release strategies.
- These insights suggest that while the average duration is relatively long, there are specific time periods, such as the mode of 334.0 days, that are more prevalent in the acquisition and addition of movies to Netflix following their original release.

[219]:

[ ]: