Task - 4

Problem Statement : Analyze and visualize sentiment patterns in social media data to understand public opinion and attitudes towards specific topics or brands.

Dataset Used : https://www.kaggle.com/datasets/shivamb/netflix-shows (https://www.kaggle.com/datasets/shivamb/netflix-shows)

About Dataset : Netflix is one of the most popular media and video streaming platforms.This tabular dataset consists of listings of all the movies and tv shows available on netflix , along with details such as ; cast,directors,ratings,etc.

In [20]:
```
pip install textblob
```

```
Defaulting to user installation because normal site-packages is not writeable
Note: you may need to restart the kernel to use updated packages.

Requirement already satisfied: textblob in c:\users\adith\appdata\roaming\pyt
hon\python311\site-packages (0.18.0.post0)
Requirement already satisfied: nltk>=3.8 in c:\programdata\anaconda3\lib\site
-packages (from textblob) (3.8.1)
Requirement already satisfied: click in c:\programdata\anaconda3\lib\site-pac
kages (from nltk>=3.8->textblob) (8.0.4)
Requirement already satisfied: joblib in c:\programdata\anaconda3\lib\site-pa
ckages (from nltk>=3.8->textblob) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in c:\programdata\anaconda3\li
b\site-packages (from nltk>=3.8->textblob) (2022.7.9)
Requirement already satisfied: tqdm in c:\programdata\anaconda3\lib\site-pack
ages (from nltk>=3.8->textblob) (4.65.0)
Requirement already satisfied: colorama in c:\programdata\anaconda3\lib\site-
packages (from click->nltk>=3.8->textblob) (0.4.6)
```

In [17]:
```
#importing necessary packages
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [13]:
```
#loading the data set
df=pd.read_csv("netflix_titles.csv")
df.head()
```

Out[13]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	du
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	Se
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	S
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	S
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	Se

In [14]:
```
df.tail()
```

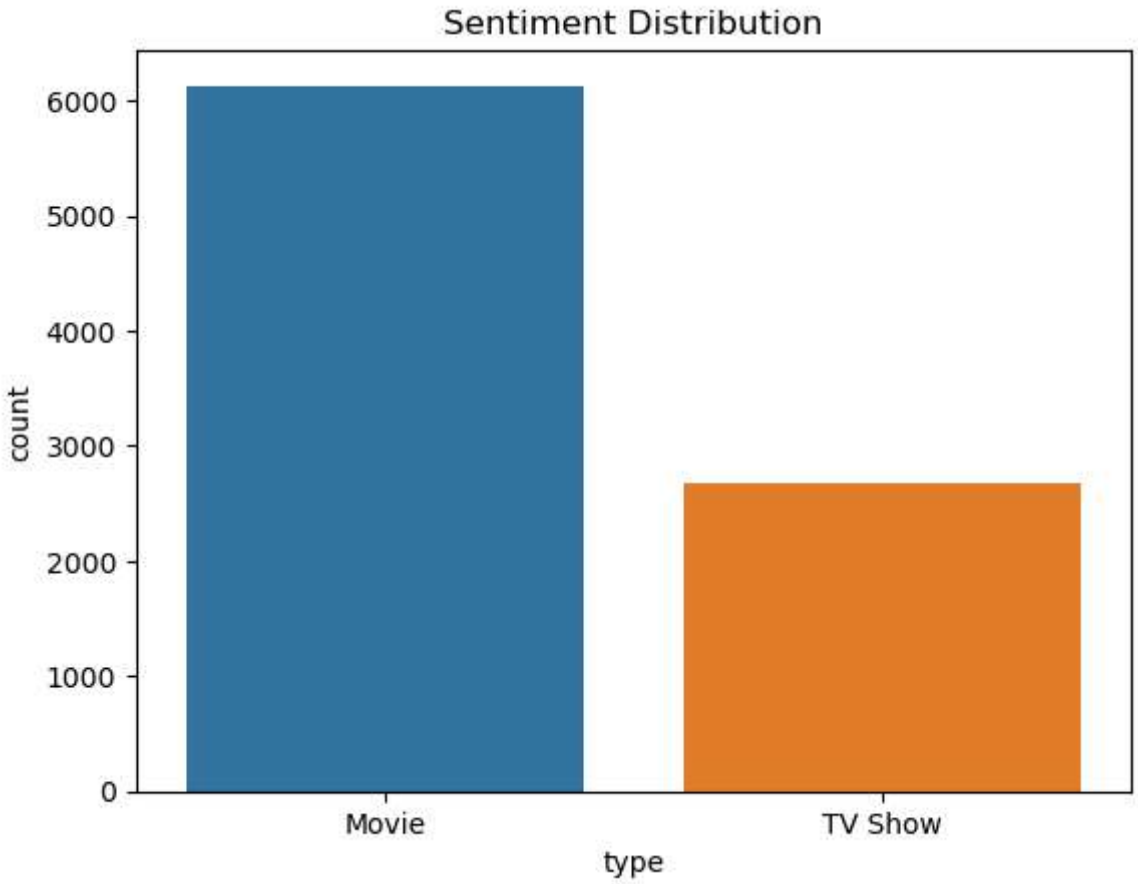
Out[14]:

	show_id	type	title	director	cast	country	date_added	release_year	ratin
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	P(
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-1

In [15]:
```
#exploring the dataset
print(df.info())
print(df['type'].value_counts())
```

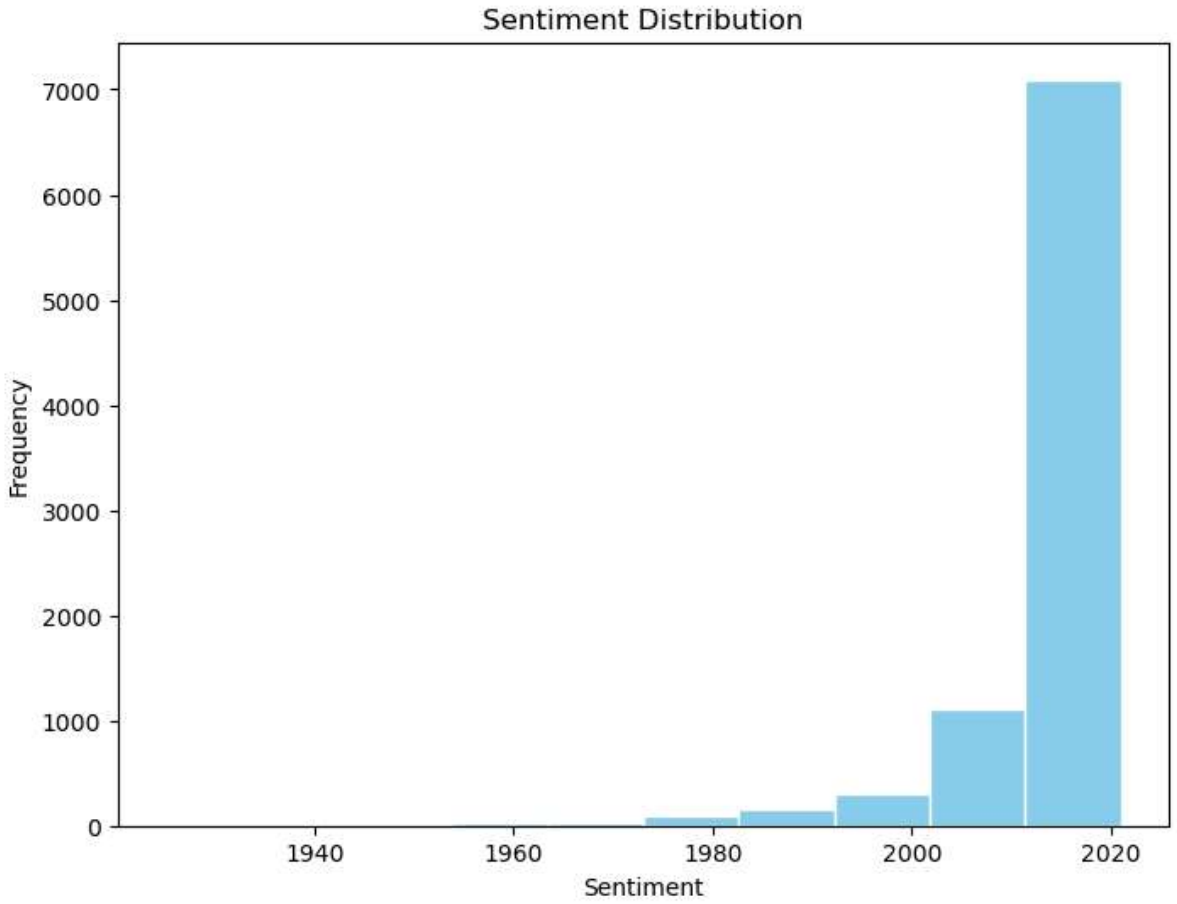
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8807 non-null   object
 1   type          8807 non-null   object
 2   title         8807 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
None
type
Movie      6131
TV Show    2676
Name: count, dtype: int64
```

In [16]:
```
#bar plot
sns.countplot(x='type',data=df)
plt.title('Sentiment Distribution')
plt.show()
```



Interpretation : It seems that Netflix releases more movies than TV shows. This observation suggests that Netflix prioritizes the production and acquisition of movies over TV shows in its content library. This strategy may be influenced by various factors, such as audience preferences, licensing agreements, production costs, and market demand. By focusing on movie releases, Netflix aims to offer a diverse range of content to its subscribers, catering to different viewing preferences and enhancing its competitive edge in the streaming industry.

In [43]:
```
plt.figure(figsize=(8,6))
plt.hist(df['release_year'],color='skyblue',edgecolor='white')
plt.xlabel('Sentiment')
plt.ylabel('Frequency')
plt.title('Sentiment Distribution')
plt.show()
```



Interpretation : The number of releases through Netflix is increasing by year, indicating a growing trend in content production and acquisition by the platform. This upward trajectory suggests that Netflix is continuously expanding its library of movies and TV shows to cater to the evolving preferences of its global audience.

```
WORD CLOUD
```

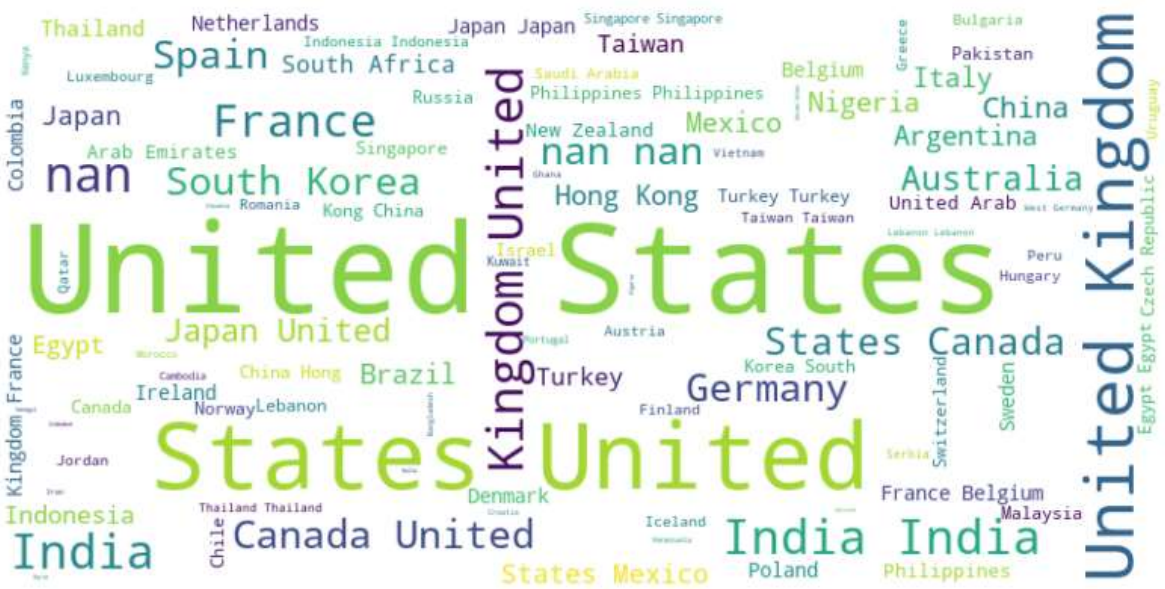
In [33]:
```
pip install wordcloud
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: wordcloud in c:\users\adith\appdata\roaming\py
thon\python311\site-packages (1.9.3)
Requirement already satisfied: numpy>=1.6.1 in c:\programdata\anaconda3\lib\s
ite-packages (from wordcloud) (1.24.3)
Requirement already satisfied: pillow in c:\programdata\anaconda3\lib\site-pa
ckages (from wordcloud) (9.4.0)
Requirement already satisfied: matplotlib in c:\programdata\anaconda3\lib\sit
e-packages (from wordcloud) (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\programdata\anaconda3\l
ib\site-packages (from matplotlib->wordcloud) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\programdata\anaconda3\lib\s
ite-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\programdata\anaconda3
\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\programdata\anaconda3
\lib\site-packages (from matplotlib->wordcloud) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\programdata\anaconda3\li
b\site-packages (from matplotlib->wordcloud) (23.1)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in c:\programdata\anacon
da3\lib\site-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\programdata\anacond
a3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\lib\site-
packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [36]:
```
from wordcloud import WordCloud
import matplotlib.pyplot as plt

text_data = ' '.join(df['country'].astype(str))
wordcloud = WordCloud(width=800, height=400, background_color='white').generat

plt.figure(figsize=(10,5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```



In [39]:
```
text_data = ' '.join(df['cast'].astype(str))
wordcloud = WordCloud(width=1000, height=700, background_color='black').genera

plt.figure(figsize=(10,5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```