In [3]:
```python
import pandas as pd
import numpy as np
```

In [4]:
```python
file_path = r"C:\Users\Aadish Sauda\login-project\Downloads\netflix.csv"
```

In [5]:
```python
df = pd.read_csv(file_path)
```

In [6]:
```python
df
```

Out[6]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
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
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA
...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy	United States	January 11, 2020	2006	PG

	show_id	type	title	director	cast	country	date_added	release_year	rating
					Chase, Kate Ma...				
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14

In [6]:
```python
columns_to_unnest = ['cast', 'listed_in']  # Add more columns if needed
for column in columns_to_unnest:
    df[column] = df[column].str.split(', ')
    df = df.explode(column)

print(df)
```

```
        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
1            s2   TV Show          Blood & Water               NaN
...         ...       ...                   ...               ...
8806      s8807     Movie                 Zubaan       Mozez Singh
8806      s8807     Movie                 Zubaan       Mozez Singh
8806      s8807     Movie                 Zubaan       Mozez Singh
8806      s8807     Movie                 Zubaan       Mozez Singh
8806      s8807     Movie                 Zubaan       Mozez Singh

                     cast        country        date_added  release_year  \
0                     NaN  United States  September 25, 2021          2020
1             Ama Qamata   South Africa  September 24, 2021          2021
1             Ama Qamata   South Africa  September 24, 2021          2021
1             Ama Qamata   South Africa  September 24, 2021          2021
1            Khosi Ngema   South Africa  September 24, 2021          2021
...                   ...            ...                ...           ...
8806       Anita Shabdish          India      March 2, 2019          2015
8806       Anita Shabdish          India      March 2, 2019          2015
8806  Chittaranjan Tripathy          India      March 2, 2019          2015
8806  Chittaranjan Tripathy          India      March 2, 2019          2015
8806  Chittaranjan Tripathy          India      March 2, 2019          2015

      rating   duration              listed_in  \
0      PG-13     90 min          Documentaries
1      TV-MA   2 Seasons  International TV Shows
1      TV-MA   2 Seasons             TV Dramas
1      TV-MA   2 Seasons           TV Mysteries
1      TV-MA   2 Seasons  International TV Shows
...      ...       ...                    ...
8806   TV-14    111 min    International Movies
8806   TV-14    111 min        Music & Musicals
8806   TV-14    111 min                 Dramas
8806   TV-14    111 min    International Movies
8806   TV-14    111 min        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...
1     After crossing paths at a party, a Cape Town t...
1     After crossing paths at a party, a Cape Town t...
1     After crossing paths at a party, a Cape Town t...
...                                           ...
8806  A scrappy but poor boy worms his way into a ty...
8806  A scrappy but poor boy worms his way into a ty...
8806  A scrappy but poor boy worms his way into a ty...
8806  A scrappy but poor boy worms his way into a ty...
8806  A scrappy but poor boy worms his way into a ty...

[149512 rows x 12 columns]
```

In [7]:
```python
categorical_variables = ['director', 'cast', 'country', 'rating']
for column in categorical_variables:
    df[column].fillna('Unknown_' + column, inplace=True)


continuous_variables = ['release_year', 'duration']
for column in continuous_variables:
    df[column].fillna(0, inplace=True)

print(df)
```

```
      show_id      type                title            director  \
0          s1     Movie  Dick Johnson Is Dead     Kirsten Johnson
1          s2   TV Show         Blood & Water   Unknown_director
1          s2   TV Show         Blood & Water   Unknown_director
1          s2   TV Show         Blood & Water   Unknown_director
1          s2   TV Show         Blood & Water   Unknown_director
...       ...       ...                   ...                ...
8806    s8807     Movie                Zubaan        Mozez Singh
8806    s8807     Movie                Zubaan        Mozez Singh
8806    s8807     Movie                Zubaan        Mozez Singh
8806    s8807     Movie                Zubaan        Mozez Singh
8806    s8807     Movie                Zubaan        Mozez Singh

                     cast        country          date_added  release_year  \
0           Unknown_cast  United States  September 25, 2021          2020
1             Ama Qamata   South Africa  September 24, 2021          2021
1             Ama Qamata   South Africa  September 24, 2021          2021
1             Ama Qamata   South Africa  September 24, 2021          2021
1            Khosi Ngema   South Africa  September 24, 2021          2021
...                  ...            ...                 ...           ...
8806       Anita Shabdish          India      March 2, 2019          2015
8806       Anita Shabdish          India      March 2, 2019          2015
8806  Chittaranjan Tripathy          India      March 2, 2019          2015
8806  Chittaranjan Tripathy          India      March 2, 2019          2015
8806  Chittaranjan Tripathy          India      March 2, 2019          2015

      rating   duration              listed_in  \
0      PG-13     90 min            Documentaries
1      TV-MA   2 Seasons  International TV Shows
1      TV-MA   2 Seasons              TV Dramas
1      TV-MA   2 Seasons            TV Mysteries
1      TV-MA   2 Seasons  International TV Shows
...      ...       ...                      ...
8806   TV-14    111 min     International Movies
8806   TV-14    111 min         Music & Musicals
8806   TV-14    111 min                   Dramas
8806   TV-14    111 min     International Movies
8806   TV-14    111 min         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...
1      After crossing paths at a party, a Cape Town t...
1      After crossing paths at a party, a Cape Town t...
1      After crossing paths at a party, a Cape Town t...
...                                                ...
8806   A scrappy but poor boy worms his way into a ty...
8806   A scrappy but poor boy worms his way into a ty...
8806   A scrappy but poor boy worms his way into a ty...
8806   A scrappy but poor boy worms his way into a ty...
8806   A scrappy but poor boy worms his way into a ty...

[149512 rows x 12 columns]
```

```python
In [8]:  categorical_variables = ['type', 'director', 'country', 'rating']
         for column in categorical_variables:
             print(f"Counts of {column}:")
             print(df[column].value_counts())
             print()
```

```
Counts of type:
type
Movie      101692
TV Show     47820
Name: count, dtype: int64

Counts of director:
director
Unknown_director                   44621
Cathy Garcia-Molina                  356
Youssef Chahine                      288
Martin Scorsese                      273
David Dhawan                         270
                                     ...
Jon Manning                            1
Manny Rodriguez, Jay Lavender          1
Denis Hennelly, Casey Suchan           1
Fabio Ock, Joana Mazzucchelli          1
Kirsten Johnson                        1
Name: count, Length: 4529, dtype: int64

Counts of country:
country
United States                                38551
India                                        19816
Unknown_country                              11145
Japan                                         6584
United Kingdom                                5180
                                               ...
Germany, United States, Sweden                   1
United States, Botswana                          1
United States, Brazil, Japan, Spain, India       1
United States, Uruguay                           1
France, New Zealand                              1
Name: count, Length: 749, dtype: int64

Counts of rating:
rating
TV-MA           56695
TV-14           38644
R               15152
TV-PG           11944
PG-13            9860
PG               5955
TV-Y7            4287
TV-G             2435
TV-Y             2407
NR               1133
G                 728
NC-17              71
Unknown_rating     67
TV-Y7-FV           66
UR                 65
74 min              1
84 min              1
66 min              1
Name: count, dtype: int64
```
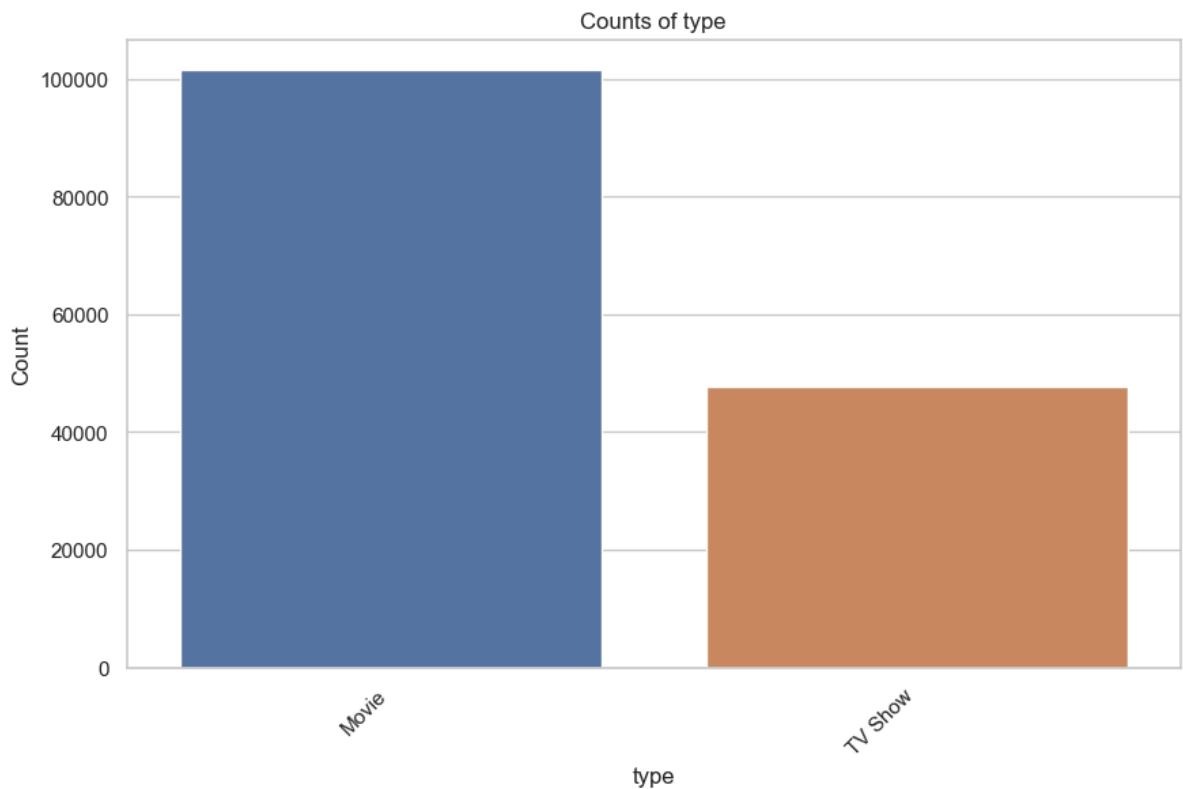
In [9]:
```python
import seaborn as sns
import matplotlib.pyplot as plt
```

In [10]:
```python
sns.set(style="whitegrid")
```

```
In [ ]:  for column in categorical_variables:
             plt.figure(figsize=(10, 6))
             sns.countplot(data=df, x=column, order=df[column].value_counts().index)
             plt.title(f"Counts of {column}")
             plt.xticks(rotation=45, ha='right')
             plt.xlabel(column)
             plt.ylabel("Count")
             plt.show()
```


Counts of type

```
In [12]:  country_counts = df[df['type'] == 'Movie'].groupby('country')['title'].nunique().re

          country_counts_sorted = country_counts.sort_values(by='title', ascending=False)

          top_10_countries = country_counts_sorted.head(10)

          print(top_10_countries)
```

```
               country  title
525      United States   2058
218              India    892
640   Unknown_country    440
440      United Kingdom   206
50              Canada    122
384               Spain    97
128               Egypt    92
319             Nigeria    86
238           Indonesia    77
428              Turkey    76
```

```
In [8]:  tv_shows_df = df[df['type'] == 'TV Show']

         tv_shows_by_country = tv_shows_df.groupby('country')['title'].nunique().reset_index

         tv_shows_by_country_sorted = tv_shows_by_country.sort_values(by='title', ascending=

         top_10_countries_tv_shows = tv_shows_by_country_sorted.head(10)

         print(top_10_countries_tv_shows)
```

```
          country  title
160    United States    760
140   United Kingdom    213
83            Japan    169
120      South Korea    158
66            India     79
132          Taiwan     68
17           Canada     59
47           France     49
4         Australia     48
125           Spain     48
```

In [7]: `print(df.head())`

```
  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        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

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

                                         listed_in  \
0                                    Documentaries
1    International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act...
3                           Docuseries, Reality TV
4  International TV Shows, Romantic TV Shows, TV ...

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

In [10]:
```python
import pandas as pd

# Read the CSV file into a DataFrame
file_path = r"C:\Users\Aadish Sauda\login-project\Downloads\netflix.csv"
df = pd.read_csv(file_path)

# Convert the 'date_added' column to datetime with multiple formats
df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')

# Drop rows with NaT values in the 'date_added' column
df.dropna(subset=['date_added'], inplace=True)

# Extract the week number from the 'date_added' column
df['week_added'] = df['date_added'].dt.isocalendar().week
```

```python
# Group the data by week and type (TV Show or Movie) and count the number of titles
weekly_counts = df.groupby(['week_added', 'type']).size().reset_index(name='count')

# Separate the counts for TV shows and movies
tv_show_counts = weekly_counts[weekly_counts['type'] == 'TV Show']
movie_counts = weekly_counts[weekly_counts['type'] == 'Movie']

# Find the best week to release TV shows
best_week_tv_show = tv_show_counts.loc[tv_show_counts['count'].idxmax()]

# Find the best week to release movies
best_week_movie = movie_counts.loc[movie_counts['count'].idxmax()]

print("Best Week to Release TV Shows:")
print(best_week_tv_show)

print("\nBest Week to Release Movies:")
print(best_week_movie)
```

```
Best Week to Release TV Shows:
week_added        27
type          TV Show
count             85
Name: 53, dtype: object

Best Week to Release Movies:
week_added         1
type           Movie
count            316
Name: 0, dtype: object
```

In [12]:
```python
import pandas as pd

# Read the CSV file into a DataFrame
file_path = r"C:\Users\Aadish Sauda\login-project\Downloads\netflix.csv"
df = pd.read_csv(file_path)

# Convert the 'date_added' column to datetime with multiple formats
df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')

# Drop rows with NaT values in the 'date_added' column
df.dropna(subset=['date_added'], inplace=True)

# Extract the month from the 'date_added' column
df['month_added'] = df['date_added'].dt.month_name()

# Group the data by month and type (TV Show or Movie) and count the number of title
monthly_counts = df.groupby(['month_added', 'type']).size().reset_index(name='count')

# Separate the counts for TV shows and movies
tv_show_counts = monthly_counts[monthly_counts['type'] == 'TV Show']
movie_counts = monthly_counts[monthly_counts['type'] == 'Movie']

# Find the best month to release TV shows
best_month_tv_show = tv_show_counts.loc[tv_show_counts['count'].idxmax()]

# Find the best month to release movies
best_month_movie = movie_counts.loc[movie_counts['count'].idxmax()]

print("Best Month to Release TV Shows:")
print(best_month_tv_show)

print("\nBest Month to Release Movies:")
print(best_month_movie)
```

```
Best Month to Release TV Shows:
month_added        July
type            TV Show
count              254
Name: 11, dtype: object

Best Month to Release Movies:
month_added        July
type              Movie
count              565
Name: 10, dtype: object
```

In [16]:
```python
import pandas as pd

file_path = r"C:\Users\Aadish Sauda\login-project\Downloads\netflix.csv"
df = pd.read_csv(file_path)

df['cast'] = df['cast'].str.split(',')

df_exploded = df.explode('cast')

actor_counts = df_exploded.groupby('cast')['title'].nunique().reset_index(name='tit

top_10_actors = actor_counts.sort_values(by='title_count', ascending=False).head(10

print("Top 10 Actors by Number of Titles Appeared In:")
print(top_10_actors)
```

```
Top 10 Actors by Number of Titles Appeared In:
                 cast  title_count
2612       Anupam Kher           39
26941     Rupa Bhimani           31
30303  Takahiro Sakurai          30
15541    Julie Tejwani           28
23624          Om Puri           27
38445   Shah Rukh Khan           26
25410     Rajesh Kava           26
33367       Yuki Kaji           25
4186      Boman Irani           25
23956     Paresh Rawal           25
```

In [15]:
```python
import pandas as pd

file_path = r"C:\Users\Aadish Sauda\login-project\Downloads\netflix.csv"
df = pd.read_csv(file_path)

df['director'] = df['director'].str.split(',')

df_exploded = df.explode('director')

director_counts = df_exploded.groupby('director')['title'].nunique().reset_index(na

director_counts = director_counts.sort_values(by='title_count', ascending=False)

top_10_directors = director_counts.head(10)

print("Top 10 Directors by Number of Titles:")
print(top_10_directors)
```

```
Top 10 Directors by Number of Titles:
               director  title_count
4020        Rajiv Chilaka           22
261            Jan Suter           18
4067         Raúl Campos           18
4651         Suhas Kadav           16
3235        Marcus Raboy           16
2450          Jay Karas            15
1382  Cathy Garcia-Molina          13
2447        Jay Chapman           12
3306      Martin Scorsese          12
5074      Youssef Chahine          12
```

In [20]:
```python
pip install wordcloud
```

```
Collecting wordcloud
  Obtaining dependency information for wordcloud from https://files.pythonhosted.o
rg/packages/f5/b0/247159f61c5d5d6647171bef84430b7efad4db504f0229674024f3a4f7f2/wor
dcloud-1.9.3-cp311-cp311-win_amd64.whl.metadata
  Downloading wordcloud-1.9.3-cp311-cp311-win_amd64.whl.metadata (3.5 kB)
Requirement already satisfied: numpy>=1.6.1 in c:\users\aadish sauda\anaconda3\lib
\site-packages (from wordcloud) (1.24.3)
Requirement already satisfied: pillow in c:\users\aadish sauda\anaconda3\lib\site-
packages (from wordcloud) (9.4.0)
Requirement already satisfied: matplotlib in c:\users\aadish sauda\anaconda3\lib\s
ite-packages (from wordcloud) (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\aadish sauda\anaconda3
\lib\site-packages (from matplotlib->wordcloud) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\users\aadish sauda\anaconda3\lib
\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\aadish sauda\anaconda
3\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\aadish sauda\anaconda
3\lib\site-packages (from matplotlib->wordcloud) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\aadish sauda\anaconda3
\lib\site-packages (from matplotlib->wordcloud) (23.1)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in c:\users\aadish sauda\anac
onda3\lib\site-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\aadish sauda\anaco
nda3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\aadish sauda\anaconda3\lib\sit
e-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)
Downloading wordcloud-1.9.3-cp311-cp311-win_amd64.whl (300 kB)
   ---------------------------------------- 0.0/300.2 kB ? eta -:--:--
   ---------------------------------------- 300.2/300.2 kB 6.2 MB/s eta 0:00:00
Installing collected packages: wordcloud
Successfully installed wordcloud-1.9.3
Note: you may need to restart the kernel to use updated packages.
```

In [22]:
```python
from wordcloud import WordCloud
```

In [24]:
```python
from wordcloud import WordCloud
import matplotlib.pyplot as plt

all_genres = ' '.join(df['listed_in'].dropna())

wordcloud = WordCloud(width=800, height=400, background_color='white').generate(all

plt.figure(figsize=(10, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.title('Word Cloud of Movie Genres')
plt.axis('off')
plt.show()
```

Word Cloud of Movie Genres

In [26]:
```python
import pandas as pd

file_path = r"C:\Users\Aadish Sauda\login-project\Downloads\netflix.csv"
df = pd.read_csv(file_path)

df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')

df = df.dropna(subset=['date_added'])

df['days_to_add'] = (df['date_added'] - pd.to_datetime(df['release_year'], format='

mode_days_to_add = df['days_to_add'].mode()[0]

print("Mode of days taken to add a movie to Netflix after release:", mode_days_to_a
```

Mode of days taken to add a movie to Netflix after release: 334 days

In [ ]: