In [1]:

```
import pandas as pd
```

In [4]:

```
pip install plotly==5.15.0
```

```
Collecting plotly==5.15.0
  Downloading plotly-5.15.0-py2.py3-none-any.whl (15.5 MB)
                                              0.0/15.5 MB ? eta -:--:--
                                              0.1/15.5 MB 2.0 MB/s eta 0:00:08
                                              0.1/15.5 MB 1.7 MB/s eta 0:00:10
                                              0.2/15.5 MB 2.1 MB/s eta 0:00:08
     -                                        0.4/15.5 MB 2.3 MB/s eta 0:00:07
     -                                        0.5/15.5 MB 2.4 MB/s eta 0:00:07
     -                                        0.7/15.5 MB 2.4 MB/s eta 0:00:07
     --                                       0.8/15.5 MB 2.5 MB/s eta 0:00:06
     --                                       1.0/15.5 MB 2.6 MB/s eta 0:00:06
     --                                       1.1/15.5 MB 2.7 MB/s eta 0:00:06
     --                                       1.1/15.5 MB 2.7 MB/s eta 0:00:06
     --                                       1.1/15.5 MB 2.7 MB/s eta 0:00:06
     --                                       1.1/15.5 MB 2.7 MB/s eta 0:00:06
     --                                       1.1/15.5 MB 2.7 MB/s eta 0:00:06
     --                                       1.1/15.5 MB 1.8 MB/s eta 0:00:08
     --                                       1.1/15.5 MB 1.8 MB/s eta 0:00:08
     --                                       1.1/15.5 MB 1.8 MB/s eta 0:00:08
     -----                                    2.1/15.5 MB 2.7 MB/s eta 0:00:05
     -----                                    2.3/15.5 MB 2.7 MB/s eta 0:00:05
     ------                                   2.4/15.5 MB 2.7 MB/s eta 0:00:05
     ------                                   2.5/15.5 MB 2.7 MB/s eta 0:00:05
     -------                                  2.7/15.5 MB 2.8 MB/s eta 0:00:05
     -------                                  2.9/15.5 MB 2.8 MB/s eta 0:00:05
     -------                                  3.0/15.5 MB 2.7 MB/s eta 0:00:05
     --------                                 3.1/15.5 MB 2.8 MB/s eta 0:00:05
     --------                                 3.2/15.5 MB 2.8 MB/s eta 0:00:05
     --------                                 3.4/15.5 MB 2.7 MB/s eta 0:00:05
     --------                                 3.5/15.5 MB 2.7 MB/s eta 0:00:05
     ---------                                3.6/15.5 MB 2.7 MB/s eta 0:00:05
     ---------                                3.7/15.5 MB 2.8 MB/s eta 0:00:05
     ---------                                3.7/15.5 MB 2.8 MB/s eta 0:00:05
     ---------                                3.7/15.5 MB 2.8 MB/s eta 0:00:05
     ---------                                3.7/15.5 MB 2.8 MB/s eta 0:00:05
     ----------                               4.2/15.5 MB 2.7 MB/s eta 0:00:05
     -----------                              4.4/15.5 MB 2.7 MB/s eta 0:00:05
     -----------                              4.5/15.5 MB 2.7 MB/s eta 0:00:05
     -----------                              4.6/15.5 MB 2.7 MB/s eta 0:00:05
     ------------                             4.7/15.5 MB 2.7 MB/s eta 0:00:04
     ------------                             4.8/15.5 MB 2.7 MB/s eta 0:00:04
     ------------                             4.9/15.5 MB 2.7 MB/s eta 0:00:04
     ------------                             5.0/15.5 MB 2.7 MB/s eta 0:00:04
     -------------                            5.2/15.5 MB 2.7 MB/s eta 0:00:04
```

```
-------------                            5.3/15.5 MB 2.7 MB/s eta 0:00:04
--------------                           5.5/15.5 MB 2.7 MB/s eta 0:00:04
--------------                           5.6/15.5 MB 2.7 MB/s eta 0:00:04
--------------                           5.8/15.5 MB 2.7 MB/s eta 0:00:04
---------------                          5.9/15.5 MB 2.7 MB/s eta 0:00:04
---------------                          6.0/15.5 MB 2.7 MB/s eta 0:00:04
---------------                          6.2/15.5 MB 2.7 MB/s eta 0:00:04
----------------                         6.3/15.5 MB 2.8 MB/s eta 0:00:04
----------------                         6.5/15.5 MB 2.8 MB/s eta 0:00:04
----------------                         6.6/15.5 MB 2.8 MB/s eta 0:00:04
-----------------                        6.7/15.5 MB 2.8 MB/s eta 0:00:04
-----------------                        6.9/15.5 MB 2.8 MB/s eta 0:00:04
------------------                       7.0/15.5 MB 2.8 MB/s eta 0:00:04
------------------                       7.1/15.5 MB 2.8 MB/s eta 0:00:04
------------------                       7.2/15.5 MB 2.7 MB/s eta 0:00:04
------------------                       7.4/15.5 MB 2.7 MB/s eta 0:00:03
-------------------                      7.5/15.5 MB 2.7 MB/s eta 0:00:03
-------------------                      7.6/15.5 MB 2.7 MB/s eta 0:00:03
-------------------                      7.7/15.5 MB 2.7 MB/s eta 0:00:03
--------------------                     7.9/15.5 MB 2.7 MB/s eta 0:00:03
--------------------                     8.0/15.5 MB 2.7 MB/s eta 0:00:03
--------------------                     8.1/15.5 MB 2.7 MB/s eta 0:00:03
---------------------                    8.2/15.5 MB 2.7 MB/s eta 0:00:03
---------------------                    8.4/15.5 MB 2.7 MB/s eta 0:00:03
---------------------                    8.5/15.5 MB 2.7 MB/s eta 0:00:03
----------------------                   8.7/15.5 MB 2.7 MB/s eta 0:00:03
----------------------                   8.8/15.5 MB 2.7 MB/s eta 0:00:03
-----------------------                  9.0/15.5 MB 2.7 MB/s eta 0:00:03
-----------------------                  9.1/15.5 MB 2.7 MB/s eta 0:00:03
-----------------------                  9.2/15.5 MB 2.7 MB/s eta 0:00:03
------------------------                 9.4/15.5 MB 2.7 MB/s eta 0:00:03
------------------------                 9.5/15.5 MB 2.7 MB/s eta 0:00:03
------------------------                 9.7/15.5 MB 2.7 MB/s eta 0:00:03
-------------------------                9.8/15.5 MB 2.7 MB/s eta 0:00:03
-------------------------                9.9/15.5 MB 2.7 MB/s eta 0:00:03
-------------------------                10.0/15.5 MB 2.7 MB/s eta 0:00:03
--------------------------               10.2/15.5 MB 2.7 MB/s eta 0:00:02
--------------------------               10.3/15.5 MB 2.7 MB/s eta 0:00:02
--------------------------               10.4/15.5 MB 2.7 MB/s eta 0:00:02
---------------------------              10.5/15.5 MB 2.7 MB/s eta 0:00:02
---------------------------              10.6/15.5 MB 2.7 MB/s eta 0:00:02
---------------------------              10.8/15.5 MB 2.7 MB/s eta 0:00:02
----------------------------             10.9/15.5 MB 2.7 MB/s eta 0:00:02
----------------------------             11.0/15.5 MB 2.7 MB/s eta 0:00:02
```

```
     ----------------------------             11.1/15.5 MB 2.7 MB/s eta 0:00:02
     -----------------------------            11.3/15.5 MB 2.7 MB/s eta 0:00:02
     -----------------------------            11.4/15.5 MB 3.0 MB/s eta 0:00:02
     -----------------------------            11.5/15.5 MB 2.9 MB/s eta 0:00:02
     -----------------------------            11.6/15.5 MB 2.9 MB/s eta 0:00:02
     ------------------------------           11.7/15.5 MB 2.9 MB/s eta 0:00:02
     ------------------------------           11.9/15.5 MB 2.8 MB/s eta 0:00:02
     -------------------------------          12.0/15.5 MB 2.8 MB/s eta 0:00:02
     -------------------------------          12.2/15.5 MB 2.8 MB/s eta 0:00:02
     -------------------------------          12.2/15.5 MB 2.8 MB/s eta 0:00:02
     -------------------------------          12.4/15.5 MB 2.7 MB/s eta 0:00:02
     --------------------------------         12.5/15.5 MB 2.7 MB/s eta 0:00:02
     --------------------------------         12.6/15.5 MB 2.7 MB/s eta 0:00:02
     --------------------------------         12.7/15.5 MB 2.7 MB/s eta 0:00:02
     ---------------------------------        12.9/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------        13.0/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------        13.2/15.5 MB 2.7 MB/s eta 0:00:01
     ----------------------------------       13.3/15.5 MB 2.7 MB/s eta 0:00:01
     ----------------------------------       13.4/15.5 MB 2.7 MB/s eta 0:00:01
     ----------------------------------       13.5/15.5 MB 2.7 MB/s eta 0:00:01
     -----------------------------------      13.6/15.5 MB 2.7 MB/s eta 0:00:01
     -----------------------------------      13.8/15.5 MB 2.7 MB/s eta 0:00:01
     ------------------------------------     14.0/15.5 MB 2.8 MB/s eta 0:00:01
     ------------------------------------     14.2/15.5 MB 2.8 MB/s eta 0:00:01
     -------------------------------------    14.4/15.5 MB 2.7 MB/s eta 0:00:01
     -------------------------------------    14.6/15.5 MB 2.7 MB/s eta 0:00:01
     --------------------------------------   14.8/15.5 MB 2.7 MB/s eta 0:00:01
     --------------------------------------   14.9/15.5 MB 2.7 MB/s eta 0:00:01
     --------------------------------------   15.1/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------------  15.2/15.5 MB 2.8 MB/s eta 0:00:01
     ---------------------------------------  15.3/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------------  15.4/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------------  15.5/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------------  15.5/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------------  15.5/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------------  15.5/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------------  15.5/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------------  15.5/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------------  15.5/15.5 MB 2.7 MB/s eta 0:00:01
     ---------------------------------------- 15.5/15.5 MB 2.5 MB/s eta 0:00:00
Requirement already satisfied: tenacity>=6.2.0 in d:\apps\anaconda\files\lib\site-packages (from plotly==5.15.0) (8.0.1)
Requirement already satisfied: packaging in d:\apps\anaconda\files\lib\site-packages (from plotly==5.15.0) (23.0)
Installing collected packages: plotly
  Attempting uninstall: plotly
```

```
    Found existing installation: plotly 5.9.0
    Uninstalling plotly-5.9.0:
      Successfully uninstalled plotly-5.9.0
Successfully installed plotly-5.15.0
Note: you may need to restart the kernel to use updated packages.
```

```
WARNING: Ignoring invalid distribution - (d:\apps\anaconda\files\lib\site-packages)
WARNING: Ignoring invalid distribution -ensorflow-intel (d:\apps\anaconda\files\lib\site-packages)
WARNING: Ignoring invalid distribution -rotobuf (d:\apps\anaconda\files\lib\site-packages)
WARNING: Ignoring invalid distribution - (d:\apps\anaconda\files\lib\site-packages)
WARNING: Ignoring invalid distribution -ensorflow-intel (d:\apps\anaconda\files\lib\site-packages)
WARNING: Ignoring invalid distribution -rotobuf (d:\apps\anaconda\files\lib\site-packages)

[notice] A new release of pip is available: 23.1.2 -> 23.2.1
[notice] To update, run: D:\apps\anaconda\files\python.exe -m pip install --upgrade pip
```

In [2]:

```python
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
import plotly.figure_factory as ff
from plotly.subplots import make_subplots
```

Read Dataset

In [3]:

```python
df = pd.read_csv("netflix_titles.csv")
df
```

Out[3]:

	h	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
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...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	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	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 ...	In a city of coaching centers known to train l...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...

	h	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...

8807 rows × 12 columns

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   h             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
```

Data Cleaning

In [5]:

```
df.nunique()
```

Out[5]:

```
h               8807
type               2
title           8804
director        4528
cast            7692
country          748
date_added      1767
release_year      74
rating            17
duration         220
listed_in        514
description     8775
dtype: int64
```

In [6]:

```
df.isna().sum()
```

Out[6]:

```
h                  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
```

In [7]:

```
df['rating'].unique()
```

Out[7]:

```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
       'TV-G', 'G', 'NC-17', '74 min', '84 min', '66 min', 'NR', nan,
       'TV-Y7-FV', 'UR'], dtype=object)
```

In [8]:

```
df[df['rating'].isna()]
```

Out[8]:

	h	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
5989	s5990	Movie	13TH: A Conversation with Oprah Winfrey & Ava ...	NaN	Oprah Winfrey, Ava DuVernay	NaN	January 26, 2017	2017	NaN	37 min	Movies	Oprah Winfrey sits down with director Ava DuVe...
6827	s6828	TV Show	Gargantia on the Verdurous Planet	NaN	Kaito Ishikawa, Hisako Kanemoto, Ai Kayano, Ka...	Japan	December 1, 2016	2013	NaN	1 Season	Anime Series, International TV Shows	After falling through a wormhole, a space-dwel...
7312	s7313	TV Show	Little Lunch	NaN	Flynn Curry, Olivia Deeble, Madison Lu, Oisín ...	Australia	February 1, 2018	2015	NaN	1 Season	Kids' TV, TV Comedies	Adopting a child's perspective, this show take...
7537	s7538	Movie	My Honor Was Loyalty	Alessandro Pepe	Leone Frisa, Paolo Vaccarino, Francesco Miglio...	Italy	March 1, 2017	2015	NaN	115 min	Dramas	Amid the chaos and horror of World War II, a c...

In [9]:

```
# drop direcor and cast columns
df.drop(columns=['director', 'cast'], inplace=True, axis=1)
df.columns
```

Out[9]:

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

In [10]:

```
df[df.date_added.isna()]
```

Out[10]:

	h	type	title	country	date_added	release_year	rating	duration	listed_in	description
6066	s6067	TV Show	A Young Doctor's Notebook and Other Stories	United Kingdom	NaN	2013	TV-MA	2 Seasons	British TV Shows, TV Comedies, TV Dramas	Set during the Russian Revolution, this comic ...
6174	s6175	TV Show	Anthony Bourdain: Parts Unknown	United States	NaN	2018	TV-PG	5 Seasons	Docuseries	This CNN original series has chef Anthony Bour...
6795	s6796	TV Show	Frasier	United States	NaN	2003	TV-PG	11 Seasons	Classic & Cult TV, TV Comedies	Frasier Crane is a snooty but lovable Seattle ...
6806	s6807	TV Show	Friends	United States	NaN	2003	TV-14	10 Seasons	Classic & Cult TV, TV Comedies	This hit sitcom follows the merry misadventure...
6901	s6902	TV Show	Gunslinger Girl	Japan	NaN	2008	TV-14	2 Seasons	Anime Series, Crime TV Shows	On the surface, the Social Welfare Agency appe...
7196	s7197	TV Show	Kikoriki	NaN	NaN	2010	TV-Y	2 Seasons	Kids' TV	A wacky rabbit and his gang of animal pals hav...
7254	s7255	TV Show	La Familia P. Luche	United States	NaN	2012	TV-14	3 Seasons	International TV Shows, Spanish-Language TV Sh...	This irreverent sitcom featues Ludovico, Feder...
7406	s7407	TV Show	Maron	United States	NaN	2016	TV-MA	4 Seasons	TV Comedies	Marc Maron stars as Marc Maron, who interviews...
7847	s7848	TV Show	Red vs. Blue	United States	NaN	2015	NR	13 Seasons	TV Action & Adventure, TV Comedies, TV Sci-Fi ...	This parody of first-person shooter games, mil...
8182	s8183	TV Show	The Adventures of Figaro Pho	Australia	NaN	2015	TV-Y7	2 Seasons	Kids' TV, TV Comedies	Imagine your worst fears, then multiply them: ...

In [11]:

```
df = df[df.date_added.notna()]
```

In [12]:

```
df.fillna(df.country.mode()[0], inplace=True)
```

In [13]:

```
df.country.isna().sum()
```

Out[13]:

```
0
```

In [14]:

```
df.isna().sum()
```

Out[14]:

```
h               0
type            0
title           0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
description     0
dtype: int64
```

In [15]:

```
df.head()
```

Out[15]:

	h	type	title	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	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	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	United States	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	United States	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	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 l...

In [16]:

```
# add new columns for month and year
df['year_added'] = df.date_added.apply(lambda x: x.split(",")[1])
df['month_added'] = df.date_added.apply(lambda x: x.split(" ")[0])
```

In [17]:

```
df
```

Out[17]:

	h	type	title	country	date_added	release_year	rating	duration	listed_in	description	year_added	month_added
0	s1	Movie	Dick Johnson Is Dead	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	2021	September
1	s2	TV Show	Blood & Water	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...	2021	September
2	s3	TV Show	Ganglands	United States	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...	2021	September
3	s4	TV Show	Jailbirds New Orleans	United States	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...	2021	September
4	s5	TV Show	Kota Factory	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...	2021	September
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...	2019	November
8803	s8804	TV Show	Zombie Dumb	United States	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...	2019	July
8804	s8805	Movie	Zombieland	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...	2019	November
8805	s8806	Movie	Zoom	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a	2020	January

	h	type	title	country	date_added	release_year	rating	duration	listed_in	description	year_added	month_added
										former superhero...		
8806	s8807	Movie	Zubaan	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	2019	March

8797 rows × 12 columns

In [18]:
```
rating_ages = {
    'TV-PG': 'Older Kids',
    'TV-MA': 'Adults',
    'TV-Y7-FV': 'Older Kids',
    'TV-Y7': 'Older Kids',
    'TV-14': 'Teens',
    'R': 'Adults',
    'TV-Y': 'Kids',
    'NR': 'Adults',
    'PG-13': 'Teens',
    'TV-G': 'Kids',
    'PG': 'Older Kids',
    'G': 'Kids',
    'UR': 'Adults',
    'NC-17': 'Adults'
}
```

In [19]:
```
df['target_ages'] = df.rating.replace(rating_ages)
df.target_ages.unique()
```

Out[19]:
```
array(['Teens', 'Adults', 'Older Kids', 'Kids', '74 min', '84 min',
       '66 min', 'United States'], dtype=object)
```

In [20]:
```
df['principal_country'] = df.country.apply(lambda x: x.split(",")[0])
df
```

Out[20]:

	h	type	title	country	date_added	release_year	rating	duration	listed_in	description	year_added	month_added	target_ages
0	s1	Movie	Dick Johnson Is Dead	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	2021	September	Teens
1	s2	TV Show	Blood & Water	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...	2021	September	Adults
2	s3	TV Show	Ganglands	United States	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...	2021	September	Adults
3	s4	TV Show	Jailbirds New Orleans	United States	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...	2021	September	Adults
4	s5	TV Show	Kota Factory	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 l...	2021	September	Adults
...	...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...	2019	November	Adults
8803	s8804	TV Show	Zombie Dumb	United States	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky	2019	July	Older Kids

town, a
young g...

	h	type	title	country	date_added	release_year	rating	duration	listed_in	description	year_added	month_added	target_ages
8804	s8805	Movie	Zombieland	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...	2019	November	Adults
8805	s8806	Movie	Zoom	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...	2020	January	Older Kids
8806	s8807	Movie	Zubaan	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...	2019	March	Teens

8797 rows × 14 columns

In [21]:
```
df.dtypes
```

Out[21]:
```
h                   object
type                object
title               object
country             object
date_added          object
release_year         int64
rating              object
duration            object
listed_in           object
description         object
year_added          object
month_added         object
target_ages         object
principal_country   object
dtype: object
```

In [23]:
```
# convert type and target_ages column type to category
df['type'] = pd.Categorical(df['type'])
df['target_ages'] = pd.Categorical(df['target_ages'], categories=['Kids', 'Older Kids', 'Teens', 'Adults'])
```

In [24]:

```
# convert year added column to numeric
df['year_added'] = pd.to_numeric(df['year_added'])
```

In [25]:

```
df.dtypes
```

Out[25]:

```
h                       object
type                  category
title                   object
country                 object
date_added              object
release_year             int64
rating                  object
duration                object
listed_in               object
description             object
year_added               int64
month_added             object
target_ages           category
principal_country       object
dtype: object
```

In [26]:

```
df.rename(columns={'h':'show_id'}, inplace=True)
```

Data Visualization

In [28]:

```
df.head()
```

Out[28]:

	show_id	type	title	country	date_added	release_year	rating	duration	listed_in	description	year_added	month_added	target_ages
0	s1	Movie	Dick Johnson Is Dead	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	2021	September	Teens
1	s2	TV Show	Blood & Water	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...	2021	September	Adults
2	s3	TV Show	Ganglands	United States	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...	2021	September	Adults
3	s4	TV Show	Jailbirds New Orleans	United States	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...	2021	September	Adults
4	s5	TV Show	Kota Factory	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...	2021	September	Adults

In [29]:

```
df['genre'] = df.listed_in.apply(lambda x: x.replace(' ,', ',').replace(', ', ','). split(','))
df.genre
```

Out[29]:

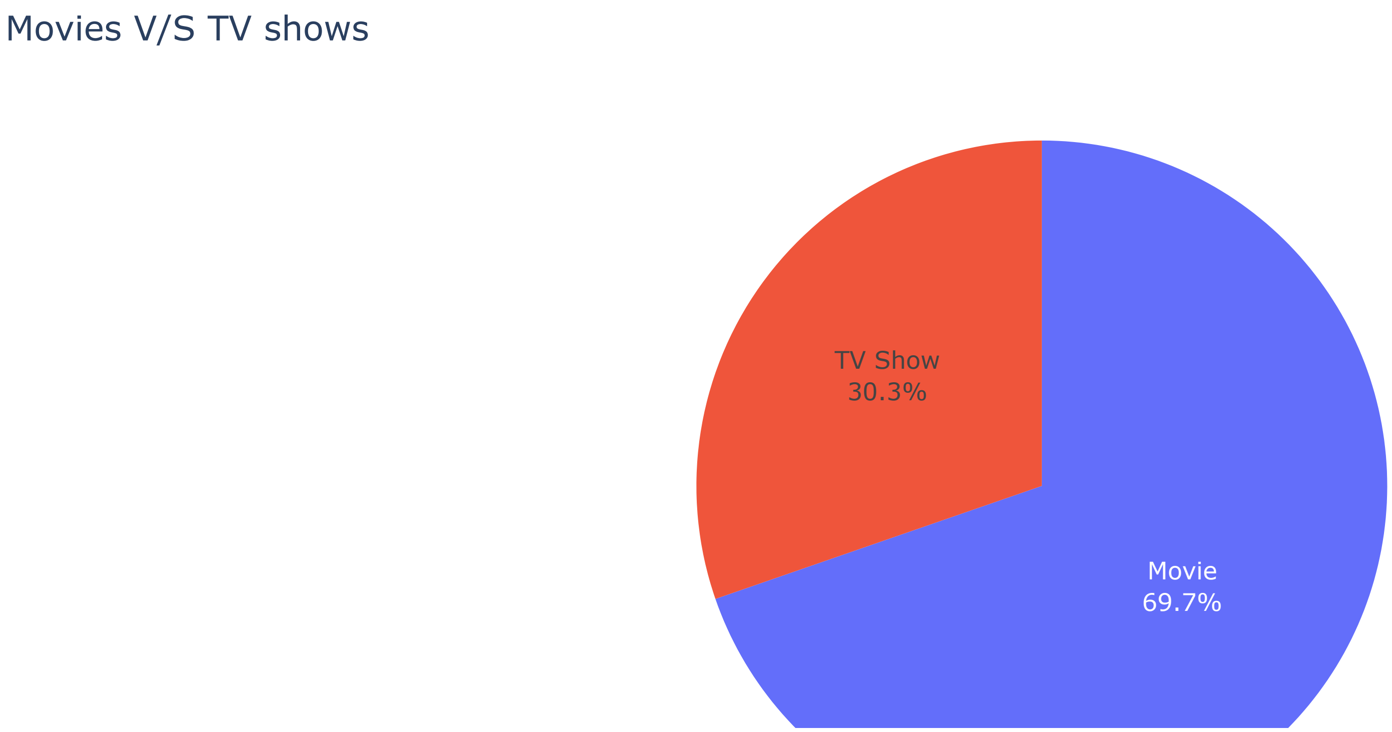
```
0                                          [Documentaries]
1        [International TV Shows, TV Dramas, TV Mysteries]
2        [Crime TV Shows, International TV Shows, TV Ac...
3                                 [Docuseries, Reality TV]
4        [International TV Shows, Romantic TV Shows, TV...
                              ...                        
8802                        [Cult Movies, Dramas, Thrillers]
8803                [Kids' TV, Korean TV Shows, TV Comedies]
8804                           [Comedies, Horror Movies]
8805                [Children & Family Movies, Comedies]
8806     [Dramas, International Movies, Music & Musicals]
Name: genre, Length: 8797, dtype: object
```

In [30]:

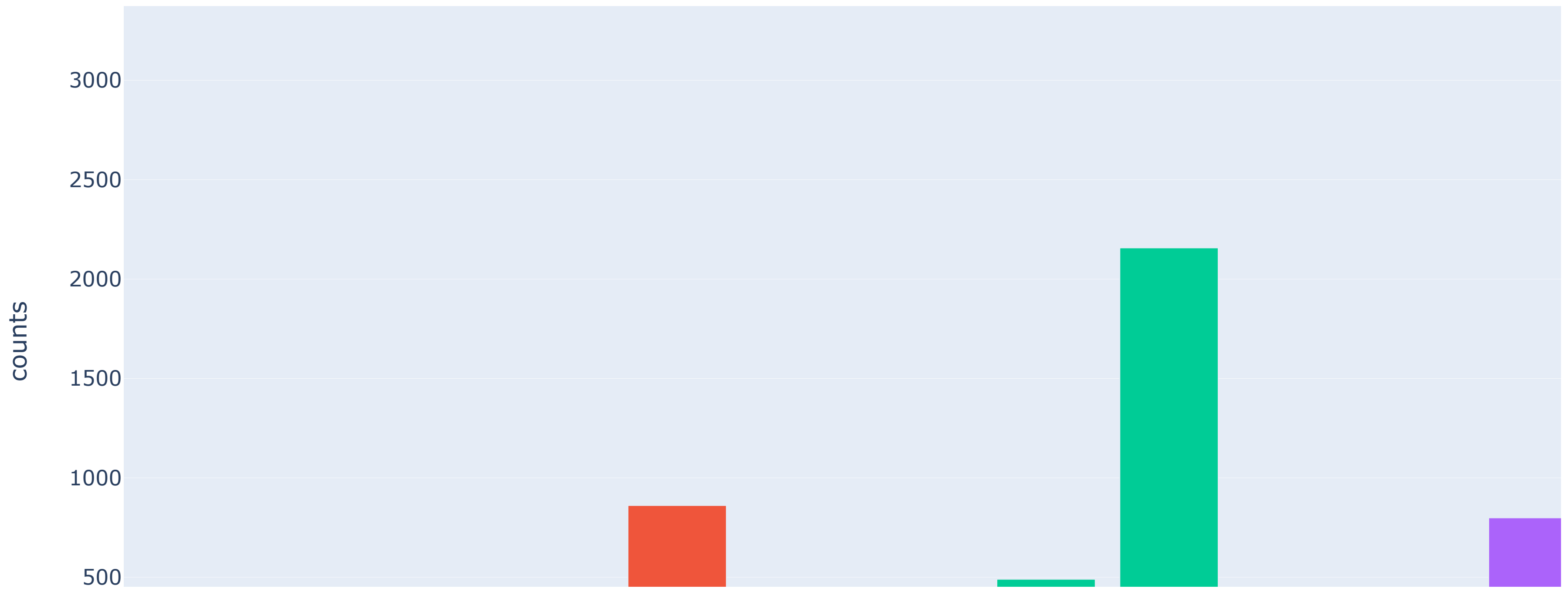
```
movie_df = df[df.type == 'Movie']
show_df = df[df.type == 'TV Show']
```

In [31]:

```
fig = px.pie(df.type.value_counts().reset_index(), values='type', names='index', title='Movies V/S TV shows')
fig.update_traces(textposition='inside', textinfo='percent+label')
fig.show()
```

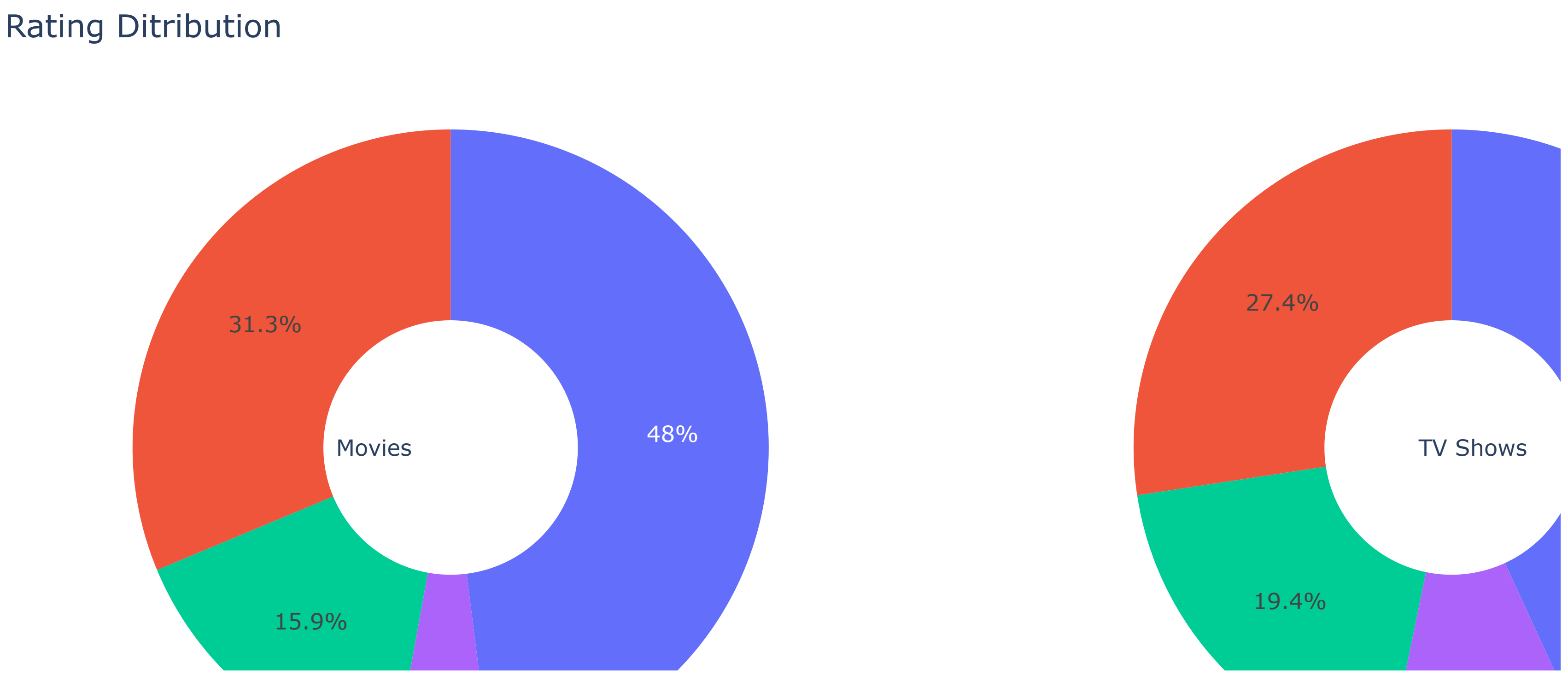


```
In [32]: rating_df = df.groupby(['rating', 'target_ages']).agg({'show_id': 'count'}).reset_index()
         rating_df = rating_df[rating_df['show_id'] != 0]
         rating_df.columns = ['rating', 'target_ages', 'counts']
         rating_df = rating_df.sort_values(by="target_ages")
         fig = px.bar(rating_df, x='rating', y='counts', color='target_ages')
         fig.show()
```



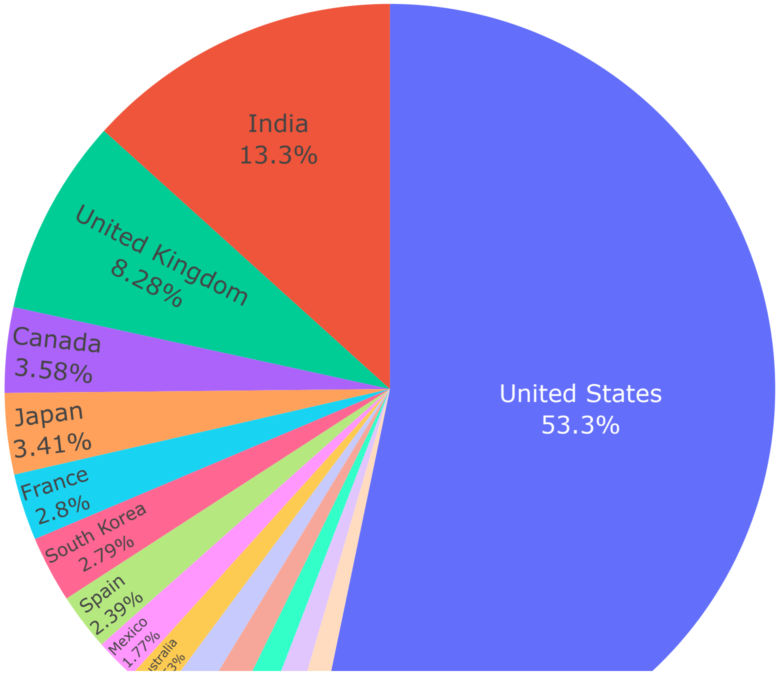
In [34]:
```
# movies rating
movie_rating_df = movie_df.groupby(['rating', 'target_ages']).agg({'show_id': 'count'}).reset_index()
movie_rating_df = movie_rating_df[movie_rating_df['show_id'] != 0]
movie_rating_df.columns = ['rating', 'target_ages', 'counts']
movie_rating_df = movie_rating_df.sort_values(by="target_ages")

# tv shows rating
show_rating_df = show_df.groupby(['rating', 'target_ages']).agg({'show_id': 'count'}).reset_index()
show_rating_df = show_rating_df[show_rating_df['show_id'] != 0]
show_rating_df.columns = ['rating', 'target_ages', 'counts']
```

```
show_rating_df = show_rating_df.sort_values(by="target_ages")


fig = make_subplots(rows=1, cols=2, specs=[[{'type': 'pie'}, {'type': 'pie'}]])

fig.add_trace(go.Pie(labels=movie_rating_df.target_ages, values=movie_rating_df.counts), row=1, col=1)
fig.add_trace(go.Pie(labels=show_rating_df.target_ages, values=show_rating_df.counts), row=1, col=2)

fig.update_traces(textposition="inside", hole=.4, hoverinfo="label+percent+name")
fig.update_layout(title_text="Rating Ditribution",
                  annotations=[dict(text='Movies', x=0.16, y=0.5, font_size=12, showarrow=False),
                               dict(text='TV Shows', x=0.82, y=0.5, font_size=12, showarrow=False)
                              ])

fig.show()
```

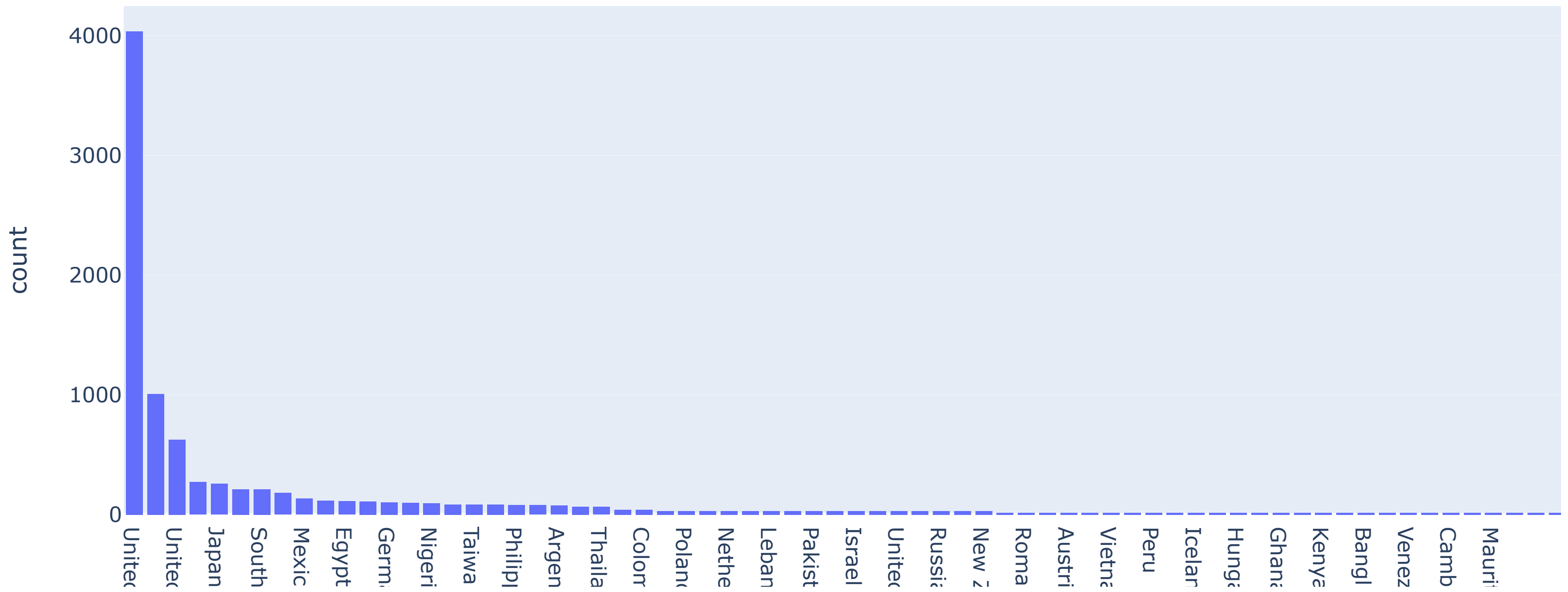


In [35]:

```
country_df = df["principal_country"].value_counts().reset_index()
country_df = country_df[country_df.principal_country/country_df.principal_country.sum() > 0.01 ]

fig = px.pie(country_df, values='principal_country', names='index')
fig.update_traces(textposition='inside', textinfo='percent+label')
fig.show()
```

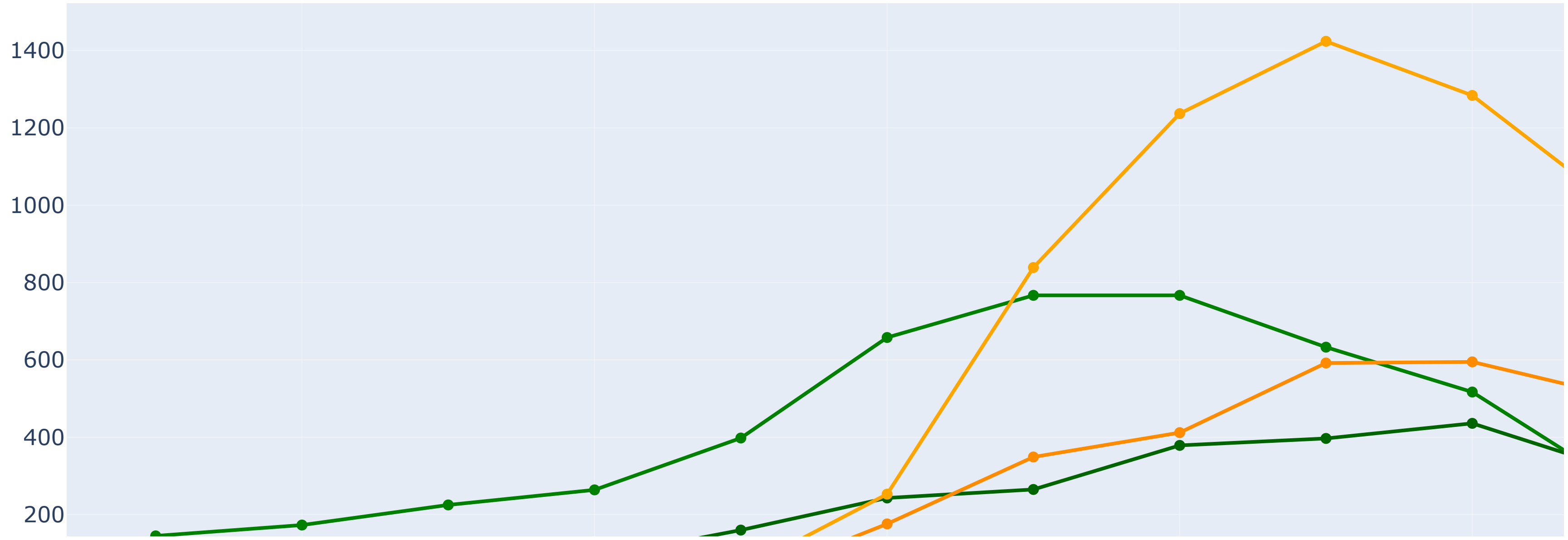


In [37]:
```
fig = px.histogram(data_frame=df, x='principal_country')
fig.update_xaxes(categoryorder='total descending')
fig.show()
```



In [54]:

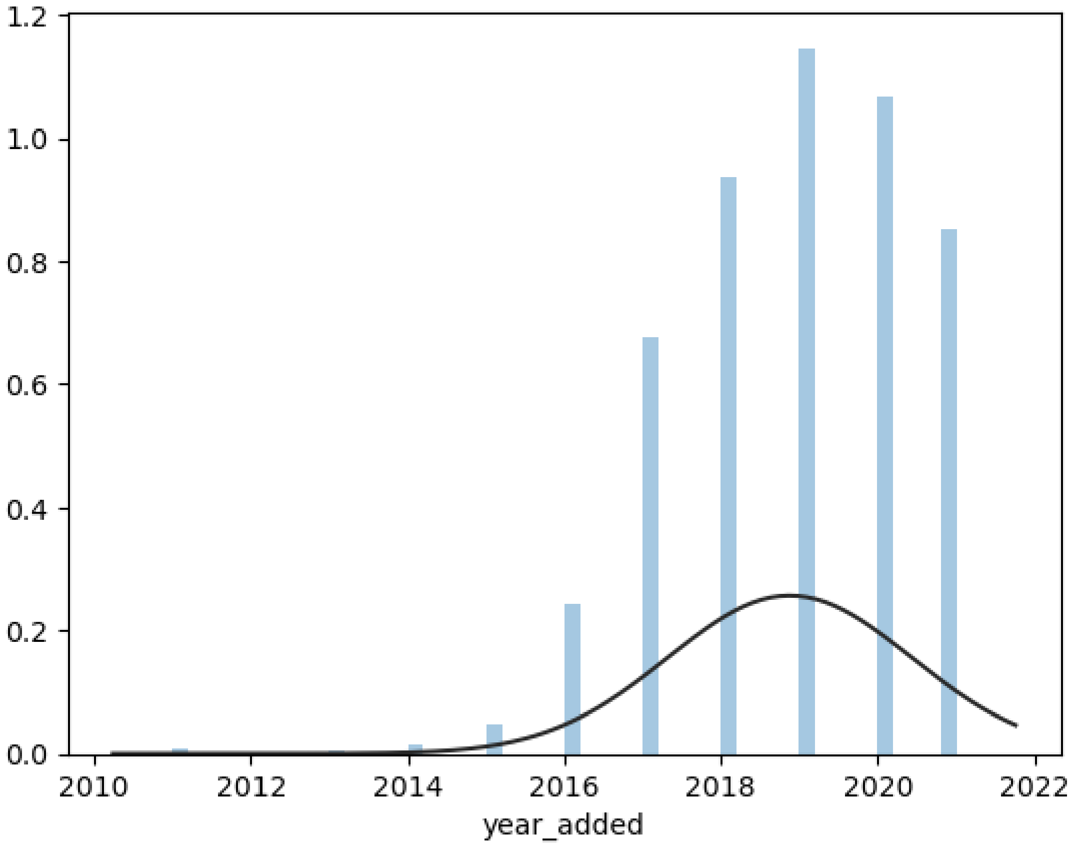
```
released_year_df = df.loc[df.release_year > 2010].groupby(["release_year", "type"]).agg({'show_id': 'count'}).reset_index()
added_year_df = df.loc[df.year_added > 2010].groupby(["year_added", "type"]).agg({'show_id': 'count'}).reset_index()

fig = go.Figure()

fig.add_trace(go.Scatter(
    x=released_year_df.loc[released_year_df['type']=='Movie']['release_year'],
    y=released_year_df.loc[released_year_df['type']=='Movie']['show_id'],
    mode='lines+markers',
    name='Movie: Released Year',
    marker_color='green'
))

fig.add_trace(go.Scatter(
    x=released_year_df.loc[released_year_df['type']=='TV Show']['release_year'],
    y=released_year_df.loc[released_year_df['type']=='TV Show']['show_id'],
    mode='lines+markers',
    name='TV Show: Released Year',
    marker_color='darkgreen'
))

fig.add_trace(go.Scatter(
    x=added_year_df.loc[added_year_df['type']=='Movie']['year_added'],
    y=added_year_df.loc[added_year_df['type']=='Movie']['show_id'],
    mode='lines+markers',
    name='Movie: Added Year',
    marker_color='orange'
))

fig.add_trace(go.Scatter(
    x=added_year_df.loc[added_year_df['type']=='TV Show']['year_added'],
    y=added_year_df.loc[added_year_df['type']=='TV Show']['show_id'],
    mode='lines+markers',
    name='TV Show: Added Year',
    marker_color='darkorange'
))

fig.show()
```



In [69]:
```
from scipy.stats import norm

sns.distplot(df.loc[df.year_added>2010, 'year_added'], fit=norm, kde=False)
```

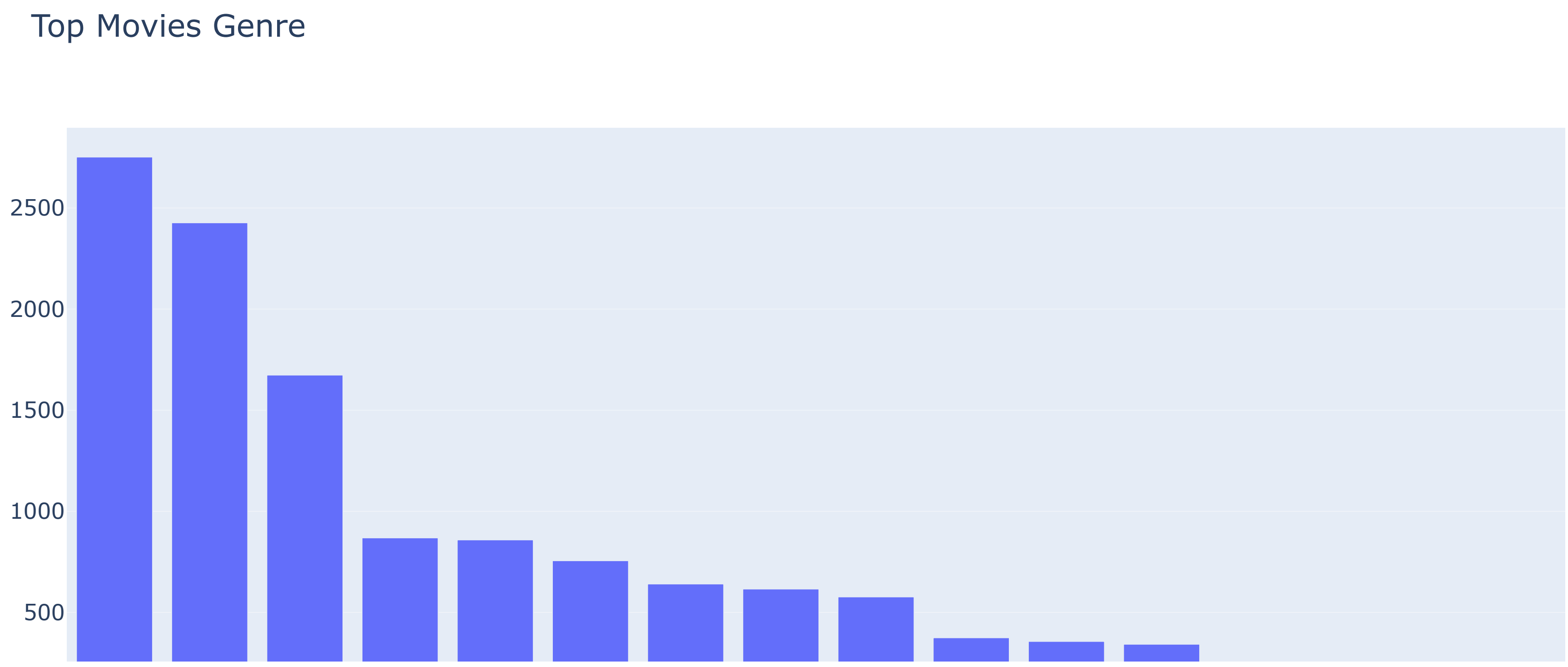
```
C:\Users\Hp\AppData\Local\Temp\ipykernel_16004\2537090449.py:3: UserWarning:



`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
```

```
Out[69]: <Axes: xlabel='year_added'>
```



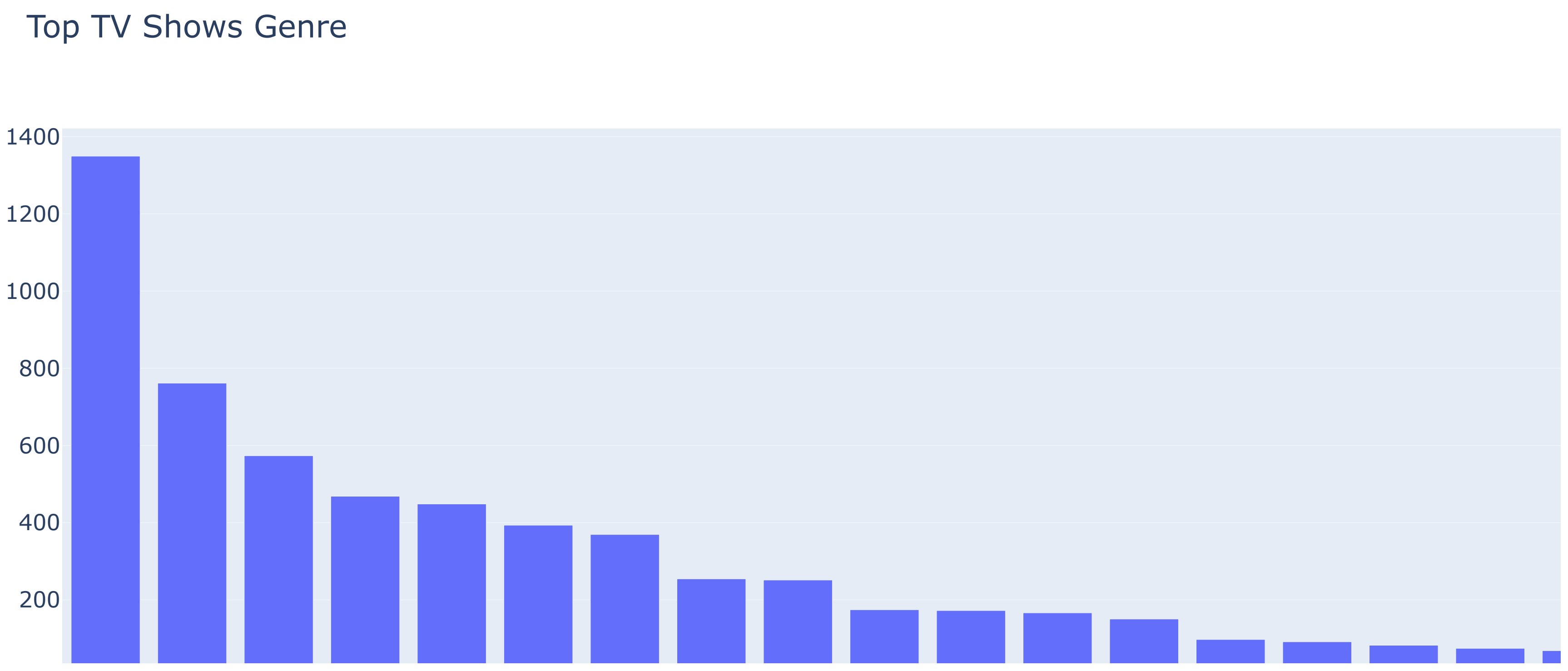
In [109…

```
from sklearn.preprocessing import MultiLabelBinarizer
mlb = MultiLabelBinarizer()
mlb_df = mlb.fit_transform(movie_df.genre)

genres_df = pd.DataFrame(mlb_df, columns=mlb.classes_, index=movie_df.index)
genres_df = genres_df.sum().sort_values(ascending=False)

fig = go.Figure()
fig.add_trace(go.Bar(x=genres_df.index, y=genres_df.values))
fig.update_layout(title="Top Movies Genre")
fig.show()
```



In [110…

```
from sklearn.preprocessing import MultiLabelBinarizer
mlb = MultiLabelBinarizer()
mlb_df = mlb.fit_transform(show_df.genre)

genres_df = pd.DataFrame(mlb_df, columns=mlb.classes_, index=show_df.index)
genres_df = genres_df.sum().sort_values(ascending=False)

fig = go.Figure()
fig.add_trace(go.Bar(x=genres_df.index, y=genres_df.values))
fig.update_layout(title="Top TV Shows Genre")
fig.show()
```

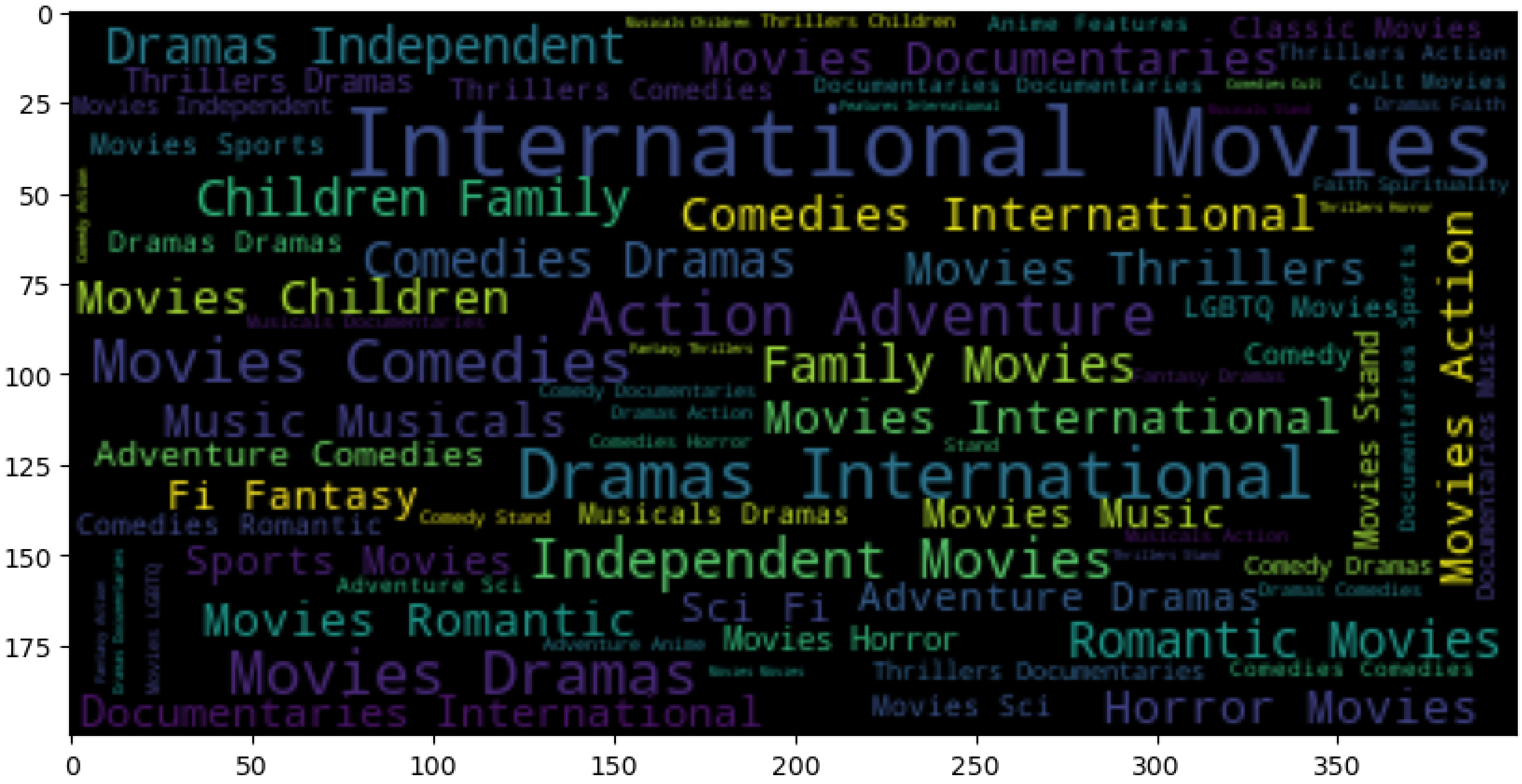


In [122…

```
from wordcloud import WordCloud
text = str(list(movie_df['genre'])).replace("[","").replace("]","").replace(",","").replace("'", "")

plt.figure(figsize=(10,10))
wordCloud = WordCloud(background_color='black').generate_from_text(text)
plt.imshow(wordCloud)
```

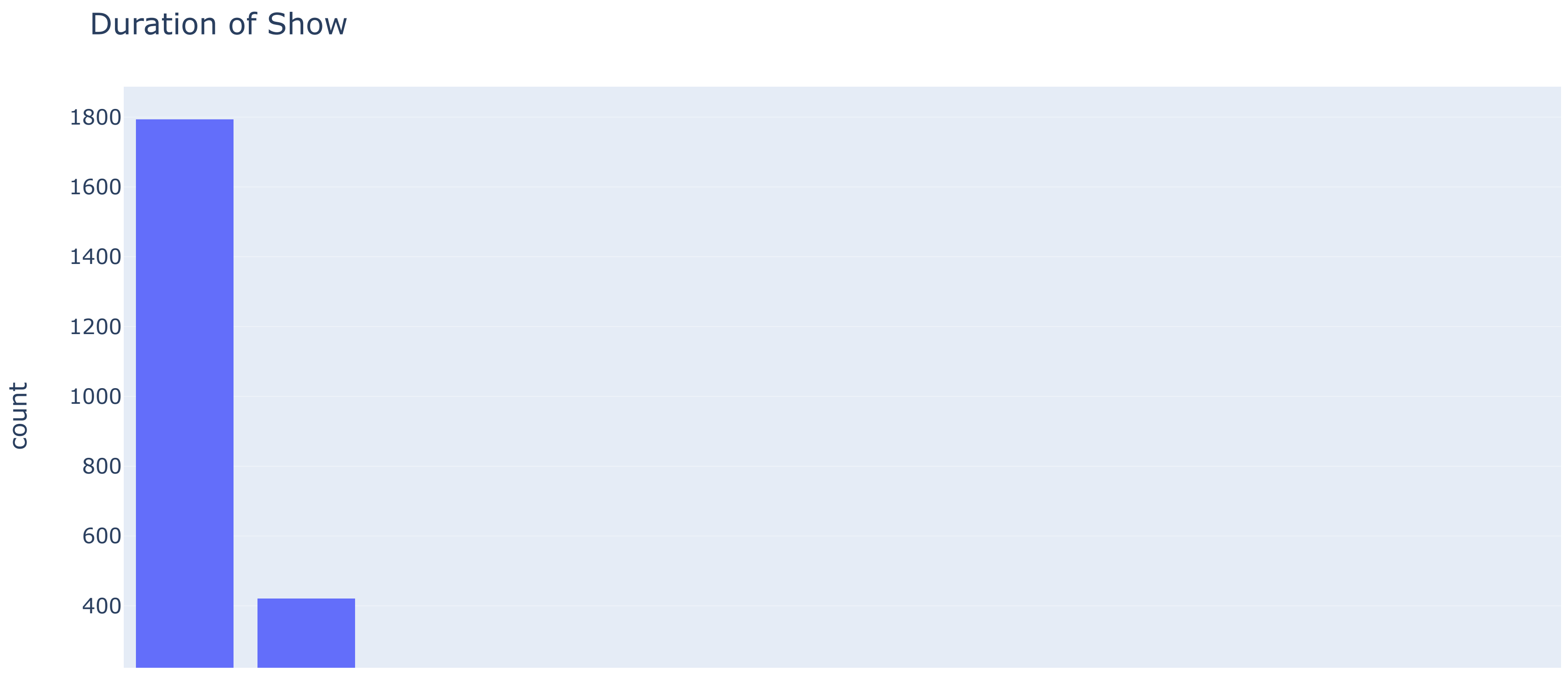
Out[122]:

```
<matplotlib.image.AxesImage at 0x25c1281b6a0>
```



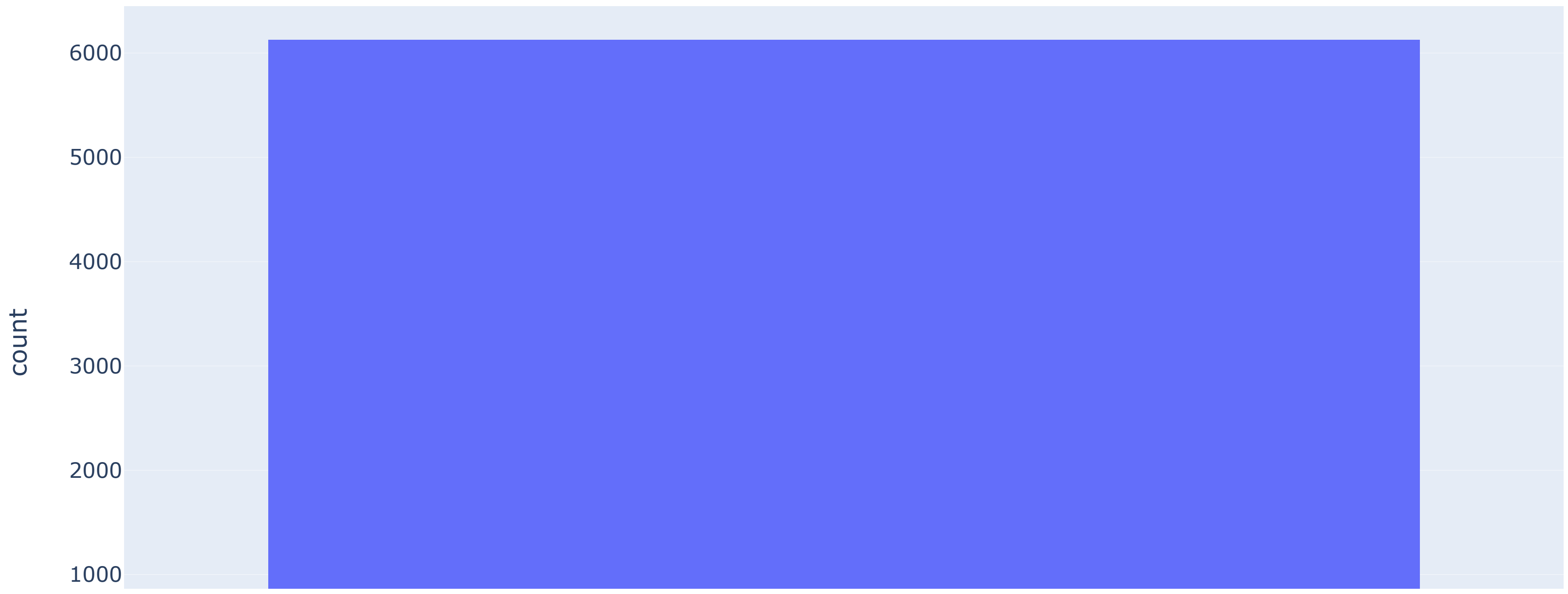
In [128…

```
fig = px.histogram(x=show_df.duration)
fig.update_xaxes(categoryorder='total descending')
fig.update_layout(title="Duration of Show")
fig.show()
```



In [171…

```
movie_df['duration_mins'] = df.duration.apply(lambda x: str(x.split(" ")[0]) if len(str(x.split(" ")[0]))>3 else 100)
movie_df = movie_df.drop(movie_df[movie_df.duration_mins=='United'].index)

movie_df.loc[movie_df.duration_mins < 90, 'duration_bin'] = 'Less than 1.30 hours'
movie_df.loc[(movie_df.duration_mins >= 90) & (movie_df.duration_mins < 150), 'duration_bin'] = 'More than 1.30 - less than 2.30
movie_df.loc[movie_df.duration_mins >= 150, 'duration_bin'] = 'More than 2.30 hours'

fig = px.histogram(x=movie_df['duration_bin'], color=movie_df['duration_bin'])
fig.show()
```



```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```