```
!gdown https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/000/940/original/netflix.csv
```

```
Downloading...
From: https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/000/940/original/netflix.csv
To: /content/netflix.csv
100% 3.40M/3.40M [00:00<00:00, 11.0MB/s]
```

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv('/content/netflix.csv')
df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descrip
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her f near end life, filr
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	cro path: party, a Tov
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 prote family fr pov drug
3	s4	TV Show	Jailbirds N...	NaN	NaN	NaN	September	2021	TV-MA	1 S	Docuseries,	F flirtation:

```python
df.columns
```

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

```python
df.info()
```

```
<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
```

```python
df.describe()
```

	release_year
count	8807.000000
mean	2014.180198
std	8.819312

```
df.describe(include = 'object')
```

	show_id	type	title	director	cast	country	date_added	rating	duration	listed_in	description
count	8807	8807	8807	6173	7982	7976	8797	8803	8804	8807	8807
unique	8807	2	8807	4528	7692	748	1767	17	220	514	8775
top	s1	Movie	Dick Johnson Is Dead	Rajiv Chilaka	David Attenborough	United States	January 1, 2020	TV-MA	1 Season	Dramas, International Movies	Paranormal activity at a lush, abandoned prope...
freq	1	6131	1	19	19	2818	109	3207	1793	362	4

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

```
show_id          0
type             0
title            0
director      2634
cast           825
```

Saved successfully!   ✕

```
rating           4
duration         3
listed_in        0
description      0
dtype: int64
```

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

```
array(['Kirsten Johnson', nan, 'Julien Leclercq', ..., 'Majid Al Ansari',
       'Peter Hewitt', 'Mozez Singh'], dtype=object)
```

```
df['director'].nunique()
```

```
4528
```

```
def uniquen(x):
  return x.unique(), x.nunique()
```

```
for i in df.columns:
  print(i, ':', uniquen(df[i]))
```

```
show_id : (array(['s1', 's2', 's3', ..., 's8805', 's8806', 's8807'], dtype=object), 8807)
type : (array(['Movie', 'TV Show'], dtype=object), 2)
title : (array(['Dick Johnson Is Dead', 'Blood & Water', 'Ganglands', ...,
       'Zombieland', 'Zoom', 'Zubaan'], dtype=object), 8807)
director : (array(['Kirsten Johnson', nan, 'Julien Leclercq', ..., 'Majid Al Ansari',
       'Peter Hewitt', 'Mozez Singh'], dtype=object), 4528)
cast : (array([nan,
       'Ama Qamata, Khosi Ngema, Gail Mabalane, Thabang Molaba, Dillon Windvogel, Natasha Thahane, Arno Greeff, Xolile Tshabalala, Ge
       'Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabiha Akkari, Sofia Lesaffre, Salim Kechiouche, Noureddine Farihi, Geert Van Rampe
       ...,
       'Jesse Eisenberg, Woody Harrelson, Emma Stone, Abigail Breslin, Amber Heard, Bill Murray, Derek Graf',
       'Tim Allen, Courteney Cox, Chevy Chase, Kate Mara, Ryan Newman, Michael Cassidy, Spencer Breslin, Rip Torn, Kevin Zegers',
       'Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary, Meghna Malik, Malkeet Rauni, Anita Shabdish, Chittaranjan
       dtype=object), 7692)
country : (array(['United States', 'South Africa', nan, 'India',
       'United States, Ghana, Burkina Faso, United Kingdom, Germany, Ethiopia',
       'United Kingdom', 'Germany, Czech Republic', 'Mexico', 'Turkey',
       'Australia', 'United States, India, France', 'Finland',
       'China, Canada, United States',
       'South Africa, United States, Japan', 'Nigeria', 'Japan',
       'Spain, United States', 'France', 'Belgium',
       'United Kingdom, United States', 'United States, United Kingdom',
       'France, United States', 'South Korea', 'Spain',
       'United States, Singapore', 'United Kingdom, Australia, France',
       'United Kingdom, Australia, France, United States',
```

```
                     'United States, Canada', 'Germany, United States',
                     'South Africa, United States', 'United States, Mexico',
                     'United States, Italy, France, Japan',
                     'United States, Italy, Romania, United Kingdom',
                     'Australia, United States', 'Argentina, Venezuela',
                     'United States, United Kingdom, Canada', 'China, Hong Kong',
                     'Russia', 'Canada', 'Hong Kong', 'United States, China, Hong Kong',
                     'Italy, United States', 'United States, Germany',
                     'United Kingdom, Canada, United States', ', South Korea',
                     'Ireland', 'India, Nepal',
                     'New Zealand, Australia, France, United States', 'Italy',
                     'Italy, Brazil, Greece', 'Argentina', 'Jordan', 'Colombia',
                     'United States, Japan', 'Belgium, United Kingdom',
                     'Switzerland, United Kingdom, Australia', 'Israel, United States',
                     'Canada, United States', 'Brazil', 'Argentina, Spain', 'Taiwan',
                     'United States, Nigeria', 'Bulgaria, United States',
                     'Spain, United Kingdom, United States', 'United States, China',
                     'United States, France',
                     'Spain, France, United Kingdom, United States',
                     ', France, Algeria', 'Poland', 'Germany',
                     'France, Israel, Germany, United States, United Kingdom',
                     'New Zealand', 'Saudi Arabia', 'Thailand', 'Indonesia',
                     'Egypt, Denmark, Germany', 'United States, Switzerland',
                     'Hong Kong, Canada, United States', 'Kuwait, United States',
                     'France, Canada, United States, Spain',
                     'France, Netherlands, Singapore', 'France, Belgium',
                     'Ireland, United States, United Kingdom', 'Egypt', 'Malaysia',
                     'Israel', 'Australia, New Zealand', 'United Kingdom, Germany',
                     'Belgium, Netherlands', 'South Korea, Czech Republic',
                     'Australia, Germany', 'Vietnam', 'United Kingdom, Belgium'
```

Saved successfully! ✕

```
df['country'].value_counts() #most type is based in United states followed by India
```

```
United States                              2818
India                                       972
United Kingdom                              419
Japan                                       245
South Korea                                 199
                                           ...
Romania, Bulgaria, Hungary                    1
Uruguay, Guatemala                            1
France, Senegal, Belgium                      1
Mexico, United States, Spain, Colombia        1
United Arab Emirates, Jordan                  1
Name: country, Length: 748, dtype: int64
```

```
df['country'].mode()
```

```
0    United States
dtype: object
```

```
#filling missing values with mode
df['country'].fillna('United States', inplace = True)
```

```
df['cast'].value_counts()
```

```
David Attenborough
19
Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jigna Bhardwaj, Rajesh Kava, Mousam, Swapnil
14
Samuel West
10
Jeff Dunham
7
David Spade, London Hughes, Fortune Feimster
6

..
Michael Peña, Diego Luna, Tenoch Huerta, Joaquin Cosio, José María Yazpik, Matt Letscher, Alyssa Diaz
1
Nick Lachey, Vanessa Lachey
1
Takeru Sato, Kasumi Arimura, Haru, Kentaro Sakaguchi, Takayuki Yamada, Kendo Kobayashi, Ken Yasuda, Arata Furuta, Suzuki Matsuo, Koichi
Yamadera, Arata Iura, Chikako Kaku, Kotaro Yoshida       1
Toyin Abraham, Sambasa Nzeribe, Chioma Chukwuka Akpotha, Chioma Omeruah, Chiwetalu Agu, Dele Odule, Femi Adebayo, Bayray McNwizu,
Biodun Stephen                                                  1
Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary, Meghna Malik, Malkeet Rauni, Anita Shabdish, Chittaranjan Tripathy
1
Name: cast, Length: 7692, dtype: int64
```

```
#filling missing values with unknown
df['cast'].fillna('Unknown', inplace = True)
```

```
#filling missing values with unknown
df['director'].fillna('Unknown', inplace = True)
```

```
df['date_added'].value_counts()
```

```
    January 1, 2020      109
    November 1, 2019      89
    March 1, 2018         75
    December 31, 2019     74
    October 1, 2018       71
                        ...
    December 4, 2016       1
    November 21, 2016      1
    November 19, 2016      1
    November 17, 2016      1
    January 11, 2020       1
    Name: date_added, Length: 1767, dtype: int64
```

```
df['date_added'].mode()
```

```
    0    January 1, 2020
    dtype: object
```

Saved successfully! ✕

```
                                          2020', inplace = True)
```

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

```
    TV-MA      3207
    TV-14      2160
    TV-PG       863
    R           799
    PG-13       490
    TV-Y7       334
    TV-Y        307
    PG          287
    TV-G        220
    NR           80
    G            41
    TV-Y7-FV      6
    NC-17         3
    UR            3
    74 min        1
    84 min        1
    66 min        1
    Name: rating, dtype: int64
```

```
df[df['rating'] == '74 min']
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
5541	s5542	Movie	Louis C.K. 2017	Louis C.K.	Louis C.K.	United States	April 4, 2017	2017	74 min	NaN	Movies	Louis C.K. muses on religion, eternal love, gi...

```
df[df['rating'] == '84 min']
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
5794	s5795	Movie	Louis C.K.: Hilarious	Louis C.K.	Louis C.K.	United States	September 16, 2016	2010	84 min	NaN	Movies	Emmy-winning comedy writer Louis C.K. brings h...

```
df[df['rating'] == '66 min']
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descriptio
5813	s5814	Movie	Louis C.K.: Live at	Louis C.K.	Louis C.K.	United States	August 15,	2015	66 min	NaN	Movies	The comic pu his tradema

```
#filling rating 74 min, 84 min and 66 min with NR(no rating)
df.loc[5541,'rating'] = 'NR'
```

```
df.loc[5794,'rating'] = 'NR'
df.loc[5813,'rating'] = 'NR'
```

```
#filling missing values with NR
df['rating'].fillna('NR', inplace = True)
```

```
df['duration'].value_counts()
```

```
    1 Season      1793
    2 Seasons      425
    3 Seasons      199
    90 min         152
    94 min         146
                   ...
    16 min           1
    186 min          1
    193 min          1
    189 min          1
    191 min          1
    Name: duration, Length: 220, dtype: int64
```

```
#imputing nan duration with the retrieved duration from rating column
df.loc[5541,'duration'] = '74 min'
df.loc[5794, 'duration'] = '84 min'
df.loc[5813, 'duration'] = '66 min'
```

Saved successfully!   ×

```
    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
```

```
df.duplicated().sum()
```

```
    0
```

```
df['type'].value_counts() # Movie is more than TV shows
```

```
    Movie      6131
    TV Show    2676
    Name: type, dtype: int64
```

```
#percentage of movies
6131/(6131+2676)
```

```
    0.6961507891449983
```

```
#percentage of TV shows
2676/(6131+2676)
```

```
    0.3038492108550017
```

```
a = df.groupby('type').type.count()
a
```

```
    type
    Movie      6131
    TV Show    2676
    Name: type, dtype: int64
```

```
plt.pie(a, labels = ['Movie', 'TV Show']) #almost 70% in netflix is movie and 30% is TV Show
```

```
([<matplotlib.patches.Wedge at 0x7f39a292ba60>,
  <matplotlib.patches.Wedge at 0x7f39a292bf10>],
 [Text(-0.6357552620136555, 0.897672126570692, 'Movie'),
  Text(0.6357552620136554, -0.8976721265706921, 'TV Show')])
```
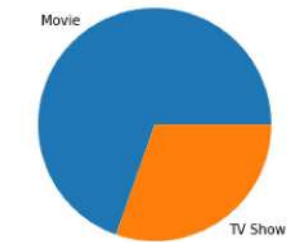


```
df1 = df.groupby(['director', 'release_year', 'date_added', 'type', 'listed_in'], as_index = False).agg({'show_id':'count'})
df1
```

	director	release_year	date_added	type	listed_in	show_id
0	A. L. Vijay	2016	May 1, 2018	Movie	Comedies, International Movies, Sci-Fi & Fantasy	1
1	A. L. Vijay	2019	September 4, 2019	Movie	Comedies, Dramas, International Movies	1
2	A. Raajdheep	2020	June 13, 2020	Movie	Dramas, International Movies	1
3	A. Salaam	1975	April 1, 2017	Movie	Action & Adventure, International Movies, Musi...	1
		2017	August 28, 2018	Movie	Action & Adventure, International Movies	1
...		...	...	...	...	...
8632	Óskar Thór Axelsson	2017	March 27, 2018	Movie	Horror Movies, International Movies	1
8633	Ömer Faruk Sorak	2004	April 23, 2021	Movie	Action & Adventure, Comedies, International Mo...	1
8634	Ömer Faruk Sorak	2011	April 4, 2018	Movie	Dramas, International Movies, Romantic Movies	1
8635	Şenol Sönmez	2015	April 23, 2021	Movie	Comedies, Dramas, International Movies	1
8636	Şenol Sönmez	2019	September 6, 2019	Movie	Comedies, International Movies, Romantic Movies	1

8637 rows × 6 columns

```
df1['show_id'].max()
```

```
6
```

```
df1[df1['show_id'] == 6] #Rajiv Chilaka has directed most movies and the genre is children & family movies in the year 2013 and in the year 2(
```

	director	release_year	date_added	type	listed_in	show_id
4490	Rajiv Chilaka	2013	July 22, 2021	Movie	Children & Family Movies	6
6606	Unknown	2016	September 19, 2016	Movie	Stand-Up Comedy	6

```
df1['show_id'].min()
```

```
1
```

```
df1[df1['show_id'] == 1] # there are 8497 records where the director has directed just a single movie or TV show per year
```

	director	release_year	date_added	type	listed_in	show_id
0	A. L. Vijay	2016	May 1, 2018	Movie	Comedies, International Movies, Sci-Fi & Fantasy	1
1	A. L. Vijay	2019	September 4, 2019	Movie	Comedies, Dramas, International Movies	1
2	A. Raaidheen	2020	June 13, 2020	Movie	Dramas, International Movies	1

```
df1['listed_in'].mode() # a lot of movies are listed in the genre Dramas and international movies
```

```
0    Dramas, International Movies
dtype: object
```

```
...          ...          ...          ...  ...                                                ...          ...
```

movies, documentaries, comedy movies are the most listed movies while action,adventure,sci-fi,sports,lgbtq, music,cult, faith and spiritualit

```
Dramas, International Movies                              361
Documentaries                                            354
Stand-Up Comedy                                          325
Comedies, Dramas, International Movies                    274
Dramas, Independent Movies, International Movies          252
                                                         ...
Comedies, LGBTQ Movies, Music & Musicals                   1
Comedies, Cult Movies, Independent Movies                  1
Dramas, Faith & Spirituality, Sports Movies                1
Action & Adventure, Classic Movies, International Movies    1
Action & Adventure, Faith & Spirituality, Sci-Fi & Fantasy 1
Name: listed_in, Length: 514, dtype: int64
```

```
[df1[(df1[....].str.contains('Kids')] #more kids TV shows can be added
```

Saved successfully! ✕

	director	release_year	date_added	type	listed_in	show_id
107	Ah Loong	2017	October 15, 2017	TV Show	Kids' TV, TV Comedies	1
361	Andrew Tan, Stephen Murray	2018	February 1, 2019	TV Show	Kids' TV	1
685	Bhavik Thakore	2017	October 18, 2020	TV Show	Kids' TV, TV Comedies	1
1405	Dheeraj Berry	2015	March 31, 2018	TV Show	Kids' TV, TV Comedies	1
1678	Fernando Moro	2020	June 19, 2021	TV Show	Kids' TV	1
...	...	...	...	...	...	...
8314	Unknown	2021	September 5, 2021	TV Show	Kids' TV, TV Comedies	1
8315	Unknown	2021	September 7, 2021	TV Show	British TV Shows, Kids' TV	1
8316	Unknown	2021	September 7, 2021	TV Show	Kids' TV, TV Comedies, TV Sci-Fi & Fantasy	1
8366	Vijay Roche	2017	May 7, 2021	TV Show	Kids' TV, TV Comedies	1
8367	Vijay S. Bhanushali	2019	March 31, 2020	TV Show	Kids' TV	1

437 rows × 6 columns

```
df1['director'].value_counts() # a lot of directors name has been unknown
```

```
Unknown                  2527
Raúl Campos, Jan Suter     17
Marcus Raboy               16
Suhas Kadav                15
Jay Karas                  14
                         ...
Jordana Spiro               1
Jordi Llompart              1
Jorge Blanco                1
Jorge Granier               1
Khaled Youssef              1
Name: director, Length: 4529, dtype: int64
```

```
df['release_year'].value_counts() #a lot of movies has been added in 2018. more no.of movies have been added after 2016
```

```
2018    1147
2017    1032
2019    1030
2020     953
2016     902
        ...
1959       1
1925       1
1961       1
```

```
    1947      1
    1966      1
    Name: release_year, Length: 74, dtype: int64
```

```
df['date_added'].value_counts() #a lot of movies have been added on January 1, 2020. a lot of movies have been added after september 2019
```

```
    January 1, 2020      119
    November 1, 2019       89
    March 1, 2018          75
    December 31, 2019      74
    October 1, 2018        71
                         ...
    December 4, 2016        1
    November 21, 2016       1
    November 19, 2016       1
    November 17, 2016       1
    January 11, 2020        1
    Name: date_added, Length: 1767, dtype: int64
```

```
df['year_added'] = df['date_added'].astype('datetime64').dt.year #to find the year the content was added in netflix
```

```
df['date_added'] = df['date_added'].astype('datetime64[ns]').dt.date #to find the date the content was added in netflix
```

```
df['day_added'] = df['date_added'].astype('datetime64[ns]').dt.day_name()#to find the day the content was added in netflix
```

Saved successfully!   ×

```
                     .astype('datetime64[ns]').dt.month_name() #to find the month the content was added in netflix
```

```
df['date'] = df['date_added'].astype('datetime64').dt.day #to find date alone
```

```
df.head()
```

|	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	year_ad
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Unknown	United States	2021-09-25	2020	PG-13	90.0	Documentaries	As her father nears the end of his life, filmm...	2
1	s2	TV Show	Blood & Water	Unknown	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV-MA	2.0	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	2
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	United States	2021-09-24	2021	TV-MA	1.0	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...	2
3	s4	TV Show	Jailbirds New Orleans	Unknown	Unknown	United States	2021-09-24	2021	TV-MA	1.0	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...	2
4	s5	TV Show	Kota Factory	Unknown	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021-09-24	2021	TV-MA	2.0	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train I...	2

```
sns.countplot(df['type'])
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword a
    warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7f3997d4ed60>
```
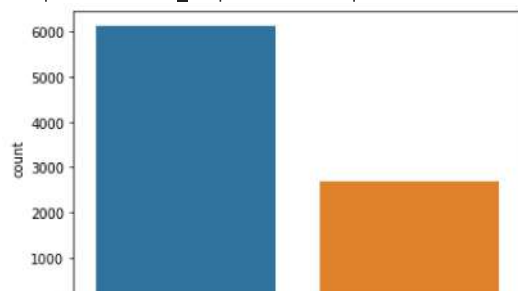


```
df['date_added'].value_counts() #max shows/movies are added in netflix on the 1st of the month and the least are been added by the end of the
```
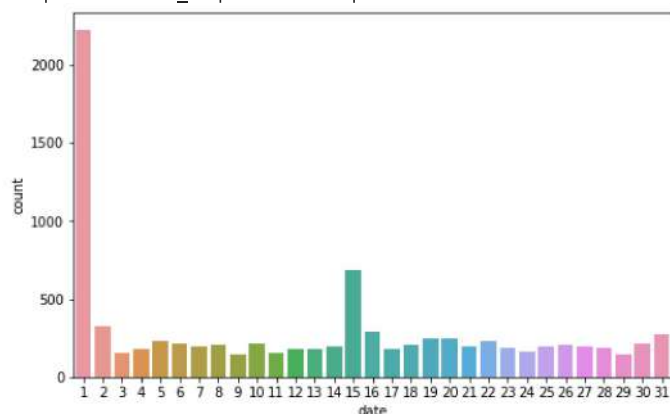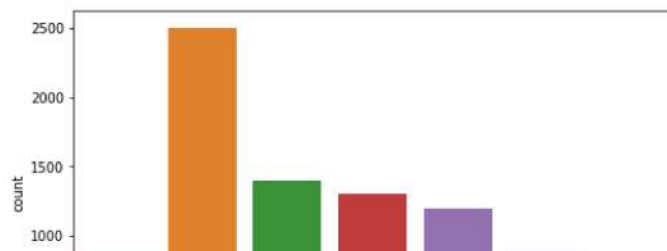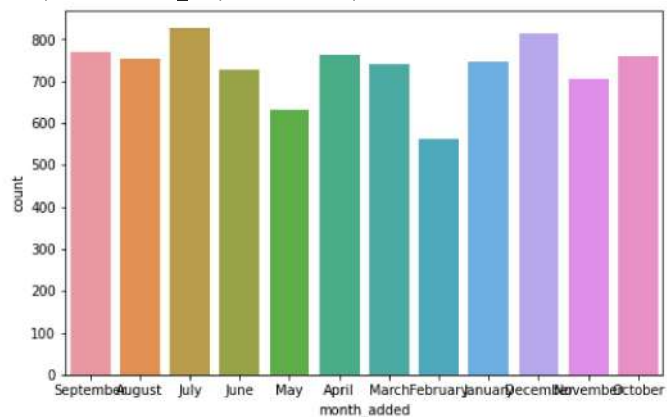
```
2020-01-01    120
2019-11-01     91
2018-03-01     75
2019-12-31     74
2018-10-01     71
              ...
2017-02-21      1
2017-02-07      1
2017-01-29      1
2017-01-25      1
2020-01-11      1
                  dtype: int64
```

Saved successfully!   ✕

```
fig = plt.figure(figsize = [8,5])
sns.countplot(df['date'])
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From v
    warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7f399624f580>
```



```
df['day_added'].value_counts() #most of the shows are added on Fridays as it is the weekend
```

```
Friday       2498
Thursday     1396
Wednesday    1298
Tuesday      1197
Monday        851
Saturday      816
Sunday        751
Name: day_added, dtype: int64
```

```
fig = plt.figure(figsize = [8,5])
sns.countplot(df['day_added'])
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From v
  warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7f3997e2b640>
```



```
df['month_added'].value_counts() #most of the shows are added on july and december
```

```
July         827
December     813
September    770
April        764
October      760
August       755
January      748
March        742
June         728
November     705
May          632
February     563
Name: month_added, dtype: int64
```

Saved successfully!

```
sns.countplot(df['month_added'])
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From v
  warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7f3997ea5b80>
```



```
df['year_added'].value_counts() #most of the shows are added in the year 2019 and the least was in the year 2010
```

```
2019    2016
2020    1889
2018    1649
2021    1498
2017    1188
2016     429
2015      82
2014      24
2011      13
2013      11
2012       3
2009       2
2008       2
2010       1
Name: year_added, dtype: int64
```

```
sns.countplot(df['year_added'])
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From
  warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7f39980f55b0>
```
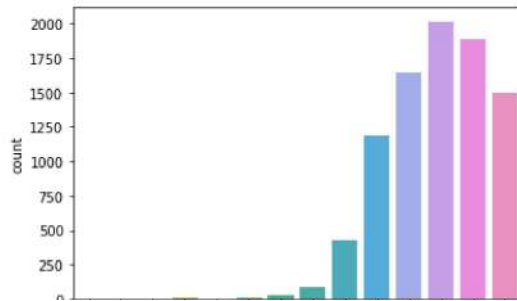


```
castlist = df['cast'].str.split(',').tolist()
castlist
```

```
[['Unknown'],
 ['Ama Qamata',
  ' Khosi Ngema',
  ' Gail Mabalane',
  ' Thabang Molaba',
  ' Dillon Windvogel',
  ' Natasha Thahane',
  ' Arno Greeff',
  ' Xolile Tshabalala',
  ' Getmore Sithole',
```

Saved successfully!   ✕

```
  ' Gretell Fincham',
  ' Sello Maake Ka-Ncube',
  ' Odwa Gwanya',
  ' Mekaila Mathys',
  ' Sandi Schultz',
  ' Duane Williams',
  ' Shamilla Miller',
  ' Patrick Mofokeng'],
 ['Sami Bouajila',
  ' Tracy Gotoas',
  ' Samuel Jouy',
  ' Nabiha Akkari',
  ' Sofia Lesaffre',
  ' Salim Kechiouche',
  ' Noureddine Farihi',
  ' Geert Van Rampelberg',
  ' Bakary Diombera'],
 ['Unknown'],
 ['Mayur More',
  ' Jitendra Kumar',
  ' Ranjan Raj',
  ' Alam Khan',
  ' Ahsaas Channa',
  ' Revathi Pillai',
  ' Urvi Singh',
  ' Arun Kumar'],
 ['Kate Siegel',
  ' Zach Gilford',
  ' Hamish Linklater',
  ' Henry Thomas',
  ' Kristin Lehman',
  ' Samantha Sloyan',
  ' Igby Rigney',
  ' Rahul Kohli',
  ' Annarah Cymone',
  ' Annabeth Gish',
  ' Alex Essoe',
  ' Rahul Abburi',
  ' Matt Biedel',
  ' Michael Trucco',
  ' Crystal Balint',
  ' Louis Oliver'],
 ['Vanessa Hudgens',
  ' Kimiko Glenn',
  ' James Marsden',
  ' Sofia Carson',
```

```
l = pd.DataFrame(pd.DataFrame(castlist).stack())
l.reset_index()
l.value_counts() #when nan/unknown actors are dropped Anupam Kher is mostly casted
```

```
 Unknown              825
  Anupam Kher           39
  Rupa Bhimani          31
  Takahiro Sakurai      30
  Julie Tejwani         28
                       ...
  João Fábio Cabral      1
  João Côrtes            1
  João Assunção          1
  Joziah Lagonoy         1
 Ṣọpẹ́ Dìrísù            1
 Length: 39297, dtype: int64
```

```
df['rating'].value_counts() # TV-MA is the mostly rated while UR and NC-17 is less
```

```
TV-MA      3207
TV-14      2160
TV-PG       863
R           799
PG-13       490
TV-Y7       334
TV-Y        307
PG          287
TV-G        220
NR           87
G            41
TV-Y7-FV      6
NC-17         3
```

Saved successfully!   ✕

```
fig = plt.figure(figsize = [8,5])
sns.countplot(df['rating'])
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From v
  warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7f399825a130>
```



```
df['duration'].value_counts() #most shows are generally added with 1 season
```

```
1 Season    1793
2 Seasons    425
3 Seasons    199
90 min       152
94 min       146
            ...
16 min         1
186 min        1
193 min        1
189 min        1
191 min        1
Name: duration, Length: 220, dtype: int64
```

```
df['duration'] = df['duration'].apply(lambda x: str(x).split()[0]).astype('float') #converting to float
```
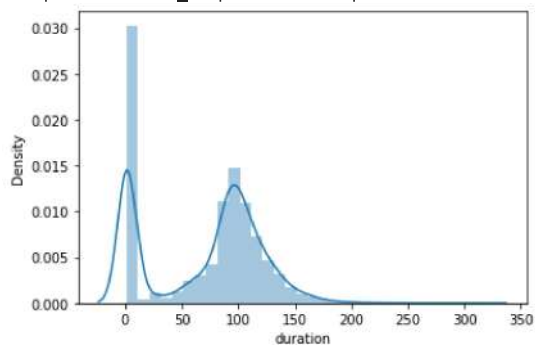
```
#distplot
sns.distplot(df['duration']) #1 season is maximum duration
```
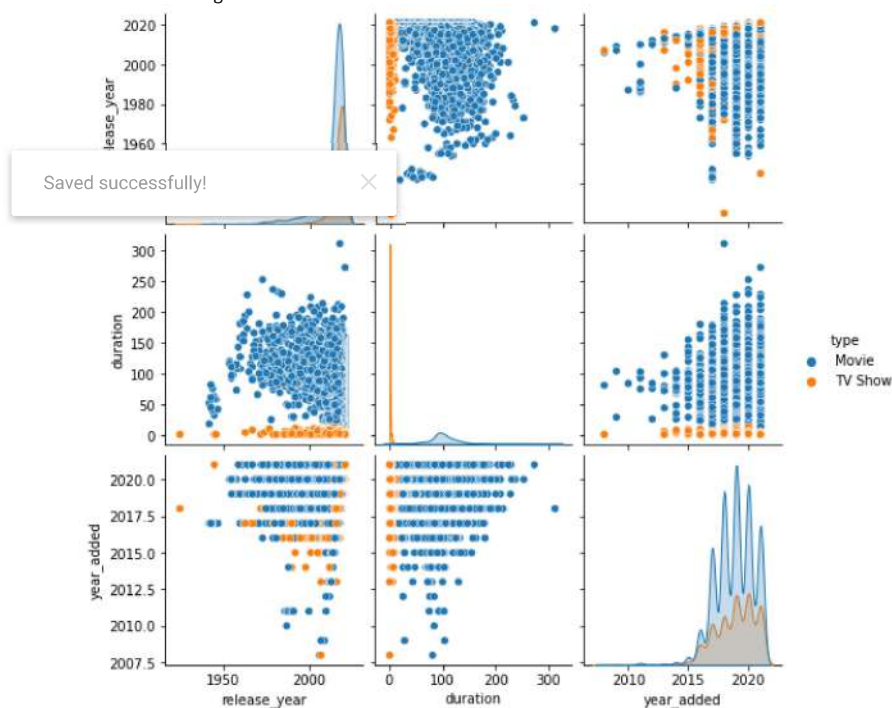
```
/usr/local/lib/python3.8/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be rem
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f39a1d446d0>
```



```
#pairplot
sns.pairplot(df, hue = 'type')
```

```
<seaborn.axisgrid.PairGrid at 0x7f39977f3640>
```



```
#Bivariate analysis
#release_year comparison for TV show and movies
fig = plt.figure(figsize = [15,5])
ax = sns.countplot(df['release_year'], hue = df['type'])
ax.set_xticklabels(ax.get_xticklabels(), rotation=90, ha="right")
plt.show()
```
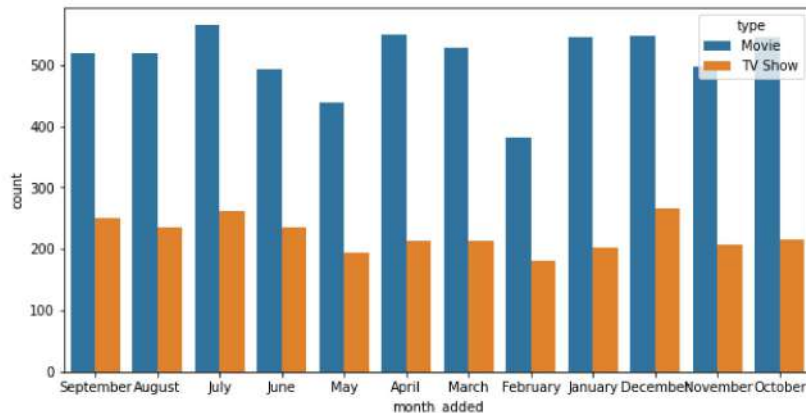
```
        /usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From v
          warnings.warn(
```

```python
#comparison of rating for types of shows
fig = plt.figure(figsize = [10,5])
sns.countplot(df['rating'], hue = df['type'])
```

```
        /usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From v
          warnings.warn(
        <matplotlib.axes._subplots.AxesSubplot at 0x7f3996488580>
```
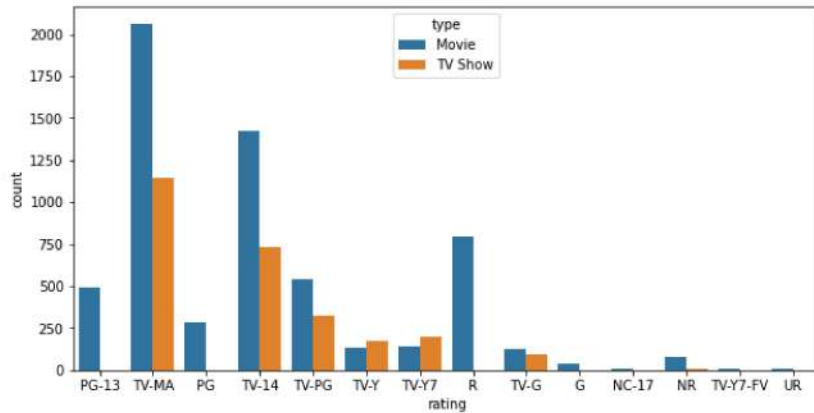

Saved successfully!   ×   of shows

```python
fig = plt.figure(figsize = [10,5])
sns.countplot(df['month_added'], hue = df['type'])
```

```
        /usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From v
          warnings.warn(
        <matplotlib.axes._subplots.AxesSubplot at 0x7f3996ed5520>
```
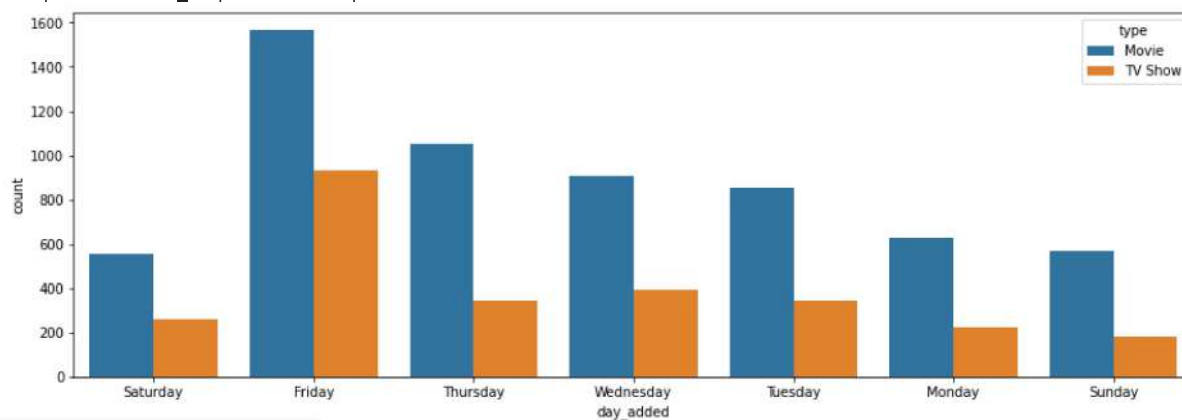


```python
#comparison of added year for types of shows
fig = plt.figure(figsize = [15,5])
sns.countplot(df['year_added'], hue = df['type'])
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From v
  warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7f399612e070>
```

```python
#comparison of added day for types of shows
fig = plt.figure(figsize = [15,5])
sns.countplot(df['day_added'], hue = df['type'])
```
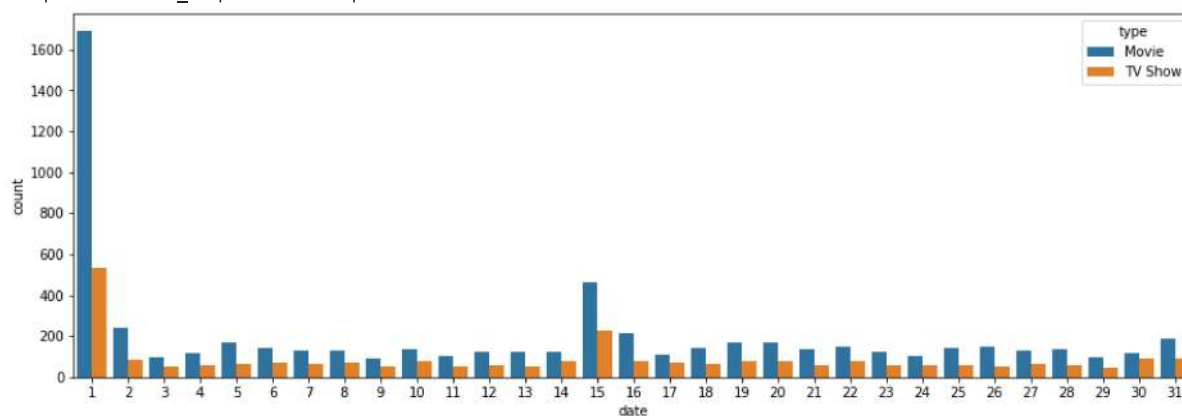
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From v
  warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7f3996614760>
```


Saved successfully!

```python
#comparison of day for types of shows
fig = plt.figure(figsize = [15,5])
sns.countplot(df['date'], hue = df['type'])
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From v
  warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7f39968c9850>
```

✓ 0s completed at 07:22

Saved successfully!