In [382]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

In [383]: | df=pd.read_csv("netflix.csv")

In [384]: | df.head()

Out[384]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	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	s 5	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 I

In [385]: df.shape

Out[385]: (8807, 12)

In [386]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):

Non-Null Count Dtype # Column 0 show_id 8807 non-null object 1 8807 non-null object type 2 title 8807 non-null object 3 director 6173 non-null object 7982 non-null 4 object cast 5 country 7976 non-null object 6 date_added 8797 non-null object 8807 non-null 7 release_year int64 8803 non-null 8 rating object 9 8804 non-null object duration listed_in 8807 non-null 10 object 11 description 8807 non-null dtypes: int64(1), object(11)

memory usage: 825.8+ KB

In [387]: df.columns

df.describe(include="object")

Out [388]:

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

In [389]: #this data set has 8807 rows and 12 columns

#there are many null values in columns:director, cast and country

#also in cast and genre colums there are many values in single row which we have to further breakdown in

In [390]: #cleaning of dataframe

In [391]: | df.head()

Out [391]: show_id type title director cast country date_added release_year rating duration listed_in description As her Dick father nears September Kirsten United PG-Johnson the end of 0 Movie NaN 2020 90 min Documentaries Johnson States 25, 2021 13 Is Dead his life, filmm... After International crossing Ama Qamata, Khosi TV Blood & South September TV-2 TV Shows, TV paths at a 2021 1 s2 NaN Ngema, Gail Show Seasons Water Africa 24, 2021 MA Dramas, TV party, a Mabalane, Thaban... Mysteries Cape Town

Julien TV-September 1 2 Gotoas, Samuel NaN 2021 International Ganglands from a Show 24, 2021 Leclercq MA Season Jouy, Nabi... TV Shows, TV powerful drug lor... Act... Feuds, flirtations **Jailbirds** TV September TV-1 Docuseries, and toilet 3 NaN NaN 2021 New NaN Show 24, 2021 MA Season Reality TV talk go Orleans down

Sami Bouajila, Tracy

In a city of International Mayur More, coaching TV Kota September TV-2 TV Shows, NaN Jitendra Kumar, India 2021 centers Show Factory 24, 2021 MΑ Seasons Romantic TV Ranjan Raj, Alam K... known to Shows, TV ... train I...

In [392]: df["title"].nunique()#we have no duplicate value in this column

Out[392]: 8807

df["title"].isna().sum()#there is no null value in this column, so this column is complteli clean.

Out[393]: 0

In [394]: |df["director"].nunique()

Out[394]: 4528

In [395]: df["director"].isna().sum()

Out[395]: 2634

In [396]: #in director column there are 2634 null values and 4528 unique which means there are many duplicate values

Crime TV

Shows,

To protect

his family

amo...

```
In [397]: df["director"].value_counts()#frequency of directors
Out[397]: Rajiv Chilaka
                                             19
          Raúl Campos, Jan Suter
                                             18
          Marcus Raboy
                                             16
          Suhas Kadav
                                             16
          Jay Karas
                                             14
          Raymie Muzquiz, Stu Livingston
                                              1
          Joe Menendez
                                              1
          Eric Bross
                                              1
          Will Eisenberg
                                              1
          Mozez Singh
          Name: director, Length: 4528, dtype: int64
In [398]: df["cast"].isna().sum()#in cast column 825 null valuses are there
Out[398]: 825
In [399]: |df["cast"].value_counts()
Out[399]: David Attenborough
          Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jigna Bhardwaj, Rajesh Kava, Mousam, Swapnil
          Samuel West
          10
          Jeff Dunham
          David Spade, London Hughes, Fortune Feimster
          Michael Peña, Diego Luna, Tenoch Huerta, Joaquin Cosio, José María Yazpik, Matt Letscher, Alyssa Diaz
          Nick Lachey, Vanessa Lachey
          Takeru Sato, Kasumi Arimura, Haru, Kentaro Sakaguchi, Takayuki Yamada, Kendo Kobayashi, Ken Yasuda, Ara
          ta Furuta, Suzuki Matsuo, Koichi Yamadera, Arata Iura, Chikako Kaku, Kotaro Yoshida
          Toyin Abraham, Sambasa Nzeribe, Chioma Chukwuka Akpotha, Chioma Omeruah, Chiwetalu Agu, Dele Odule, Fem
          i Adebayo, Bayray McNwizu, Biodun Stephen
          Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary, Meghna Malik, Malkeet Rauni, Anita S
          habdish, Chittaranjan Tripathy
          Name: cast, Length: 7692, dtype: int64
In [400]: df["country"].isna().sum()#there are 831 null values in country column
Out[400]: 831
In [401]: |df["country"].value_counts()
Out[401]: 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
          Name: country, Length: 748, dtype: int64
In [402]: #shows from united states have highest number on netflix
In [403]: df["date_added"].isna().sum()
Out [403]: 10
In [404]: df["release_year"].isna().sum()
```

Out[404]: 0

```
In [405]: | df["rating"].isna().sum()
Out[405]: 4
In [406]: |df["duration"].isna().sum()
Out[406]: 3
In [407]: |df["listed_in"].isna().sum()
Out [407]: 0
In [408]: |df["rating"].value_counts()
Out [408]: TV-MA
                       3207
          TV-14
                       2160
          TV-PG
                        863
                        799
          R
          PG-13
                        490
          TV-Y7
                        334
          TV-Y
                        307
                        287
          PG
          TV-G
                        220
          NR
                         80
          G
                         41
          TV-Y7-FV
                          6
          NC-17
                          3
          UR
                          3
          74 min
                          1
                          1
          84 min
          66 min
                          1
          Name: rating, dtype: int64
In [409]: |df["rating"] = df["rating"].replace({"74 min" : np.nan, "84 min" : np.nan, "66 min": np.nan})
          df['rating'].unique()
Out[409]: array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
                  'TV-G', 'G', 'NC-17', nan, 'NR', 'TV-Y7-FV', 'UR'], dtype=object)
In [410]: #changed the unusaual rating type into nan which will further be replaced to mode by fillna function
  In [ ]:
In [411]: |df["duration"].value_counts()
Out[411]: 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
In [412]: df["type"].isna().sum()
Out[412]: 0
In [413]: df["type"].value_counts()
Out[413]: Movie
                      6131
          TV Show
                      2676
          Name: type, dtype: int64
In [414]: #filling null values
```

```
In [415]: df["director"].fillna('Unkown_director', inplace= True)
    df["cast"].fillna('Unkown_cast', inplace= True)
    df["country"].fillna("unknown_country",inplace= True)
    df["duration"].fillna(df["duration"].mode()[0],inplace=True)
    df["date_added"].fillna(df["date_added"].mode()[0],inplace=True)
    df["rating"].fillna(df["rating"].mode()[0],inplace=True)
```

In [416]: df.head()

Out[416]:	show_		type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	des
	0	s 1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Unkown_cast	United States	September 25, 2021	2020	PG- 13	90 min	Documentaries	fath th
	1	s2	TV Show	Blood & Water	Unkown_director	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban	South Africa	September 24, 2021	2021	TV- MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	p; Ca
	2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi	unknown_country	September 24, 2021	2021	TV- MA	1 Season	Crime TV Shows, International TV Shows, TV Act	Tc h
	3	s4	TV Show	Jailbirds New Orleans	Unkown_director	Unkown_cast	unknown_country	September 24, 2021	2021	TV- MA	1 Season	Docuseries, Reality TV	f a
	4	s 5	TV Show	Kota Factory	Unkown_director	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K	India	September 24, 2021	2021	TV- MA	2 Seasons	International TV Shows, Romantic TV Shows, TV	ln c k

In [417]: #now all null values have been removed.

In [418]: #now we will do un nesting

In [419]: df["cast"] = df["cast"].apply(lambda x: str(x).split(', '))
df.head()

Out [419]: show id type title director cast country date_added release_year rating duration listed in de Dick fat September PG-2020 0 s1 Movie Johnson Kirsten Johnson [Unkown_cast] United States 90 min Documentaries 25, 2021 Is Dead [Ama Qamata, International Khosi Ngema, TV-2 TV Shows, TV Blood & September Unkown_director South Africa 2021 Gail Show Water 24, 2021 MA Seasons Dramas, TV Mabalane, Mysteries Thaba... [Sami Crime TV Bouajila, Tracy Shows, TV-September Ganglands Julien Leclercq 2 Gotoas, unknown_country 2021 International 24, 2021 MA Season TV Shows, TV Samuel Jouy, Nab... Act... Jailbirds September TV TV-Docuseries, 1 3 New Unkown_director [Unkown_cast] unknown_country 2021 Show 24, 2021 MΑ Season Reality TV Orleans [Mayur More, International Jitendra TV Shows, September TV-2 Kota Unkown_director 2021 s5 India Show Seasons Romantic TV Kumar, Ranjan 24, 2021 MΑ Factory Raj, Alam ... Shows, TV ...

In [420]: df = df.explode("cast")
df.head()

Out[420]:		show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
	0	s 1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Unkown_cast	United States	September 25, 2021	2020	PG- 13	90 min	Documentaries	As her father nears the end of his life, filmm
	1	s2	TV Show	Blood & Water	Unkown_director	Ama Qamata	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
	1	s2	TV Show	Blood & Water	Unkown_director	Khosi Ngema	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
	1	s2	TV Show	Blood & Water	Unkown_director	Gail Mabalane	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
	1	s 2	TV Show	Blood & Water	Unkown_director	Thabang Molaba	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
	df =			= df[" e("liste	listed_in"].a d_in")	apply(lambo	la x: s	tr(x).spli	it(' <mark>, '</mark>))				
Out [421]:		show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
	0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Unkown_cast	United States	September 25, 2021	2020	PG- 13	90 min	Documentaries	As her father nears the end of his life, filmm
	1	s2	TV Show	Blood & Water	Unkown_director	Ama Qamata	South Africa	September 24, 2021	2021	TV- MA	2 Seasons	International TV Shows	After crossing paths at a party, a Cape Town t
	1	s2	TV Show	Blood & Water	Unkown_director	Ama Qamata	South Africa	September 24, 2021	2021	TV- MA	2 Seasons	TV Dramas	After crossing paths at a party, a Cape Town t
	1	s2	TV Show	Blood & Water	Unkown_director	Ama Qamata	South Africa	September 24, 2021	2021	TV- MA	2 Seasons	TV Mysteries	After crossing paths at a party, a Cape Town t
						10	South	September 24, 2021	2021	TV-	2	International	After crossing paths at a
	1	s2	TV Show	Blood & Water	Unkown_director	Khosi Ngema	Africa	24, 2021		MA	Seasons	TV Shows	party, a Cape Town t
	df['	'count	Show ry"] =	Water	untry"].apply	-			, '))	MA	Seasons	TV Snows	Cape Town

```
In [423]:
```

#Now the data has been un-nested. #also we will not do anything to description column as it is of no use to us

In [424]: #breaking of date part

```
In [425]:
    df["date_added"]=pd.to_datetime(df["date_added"])
    df["week_no"]=df['date_added'].dt.week
    df["month"] = df["date_added"].dt.strftime('%b')
```

df["year"] = df["date_added"].dt.year

/var/folders/63/h28070vs2zx_1xl7yyzsblth0000gn/T/ipykernel_48797/4212171577.py:2: FutureWarning: Series.dt.weekofyear and Series.dt.week have been deprecated. Please use Series.dt.isocalendar().week instead.

df["week_no"]=df['date_added'].dt.week

In [426]: df.head()

0u

W	description	listed_in	duration	rating	release_year	date_added	country	cast	director	title	type	show_id		ut[426]:
	As her father nears the end of his life, filmm	Documentaries	90 min	PG- 13	2020	2021-09-25	United States	Unkown_cast	Kirsten Johnson	Dick Johnson Is Dead	Movie	s1	0	
	After crossing paths at a party, a Cape Town t	International TV Shows	2 Seasons	TV- MA	2021	2021-09-24	South Africa	Ama Qamata	Unkown_director	Blood & Water	TV Show	s2	1	
	After crossing paths at a party, a Cape Town t	TV Dramas	2 Seasons	TV- MA	2021	2021-09-24	South Africa	Ama Qamata	Unkown_director	Blood & Water	TV Show	s2	1	
	After crossing paths at a party, a Cape Town t	TV Mysteries	2 Seasons	TV- MA	2021	2021-09-24	South Africa	Ama Qamata	Unkown_director	Blood & Water	TV Show	s2	1	
	After crossing paths at a party, a Cape Town t	International TV Shows	2 Seasons	TV- MA	2021	2021-09-24	South Africa	Khosi Ngema	Unkown_director	Blood & Water	TV Show	s2	1	

In [427]: #dropping unnecessary columns

In [428]: | df.drop(["show_id","description"],axis=1, inplace= True)

In [429]: df.head()

Out [429]:

:		type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	week_no	month	year
•	0 1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Unkown_cast	United States	2021-09-25	2020	PG- 13	90 min	Documentaries	38	Sep	2021
	1	TV Show	Blood & Water	Unkown_director	Ama Qamata	South Africa	2021-09-24	2021	TV- MA	2 Seasons	International TV Shows	38	Sep	2021
	1	TV Show	Blood & Water	Unkown_director	Ama Qamata	South Africa	2021-09-24	2021	TV- MA	2 Seasons	TV Dramas	38	Sep	2021
	1	TV Show	Blood & Water	Unkown_director	Ama Qamata	South Africa	2021-09-24	2021	TV- MA	2 Seasons	TV Mysteries	38	Sep	2021
	1	TV Show	Blood & Water	Unkown_director	Khosi Ngema	South Africa	2021-09-24	2021	TV- MA	2 Seasons	International TV Shows	38	Sep	2021

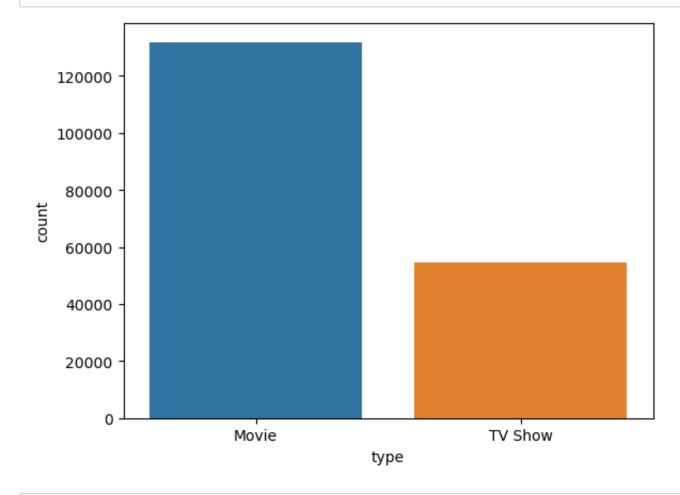
02/08/23, 9:09 PM Untitled - Jupyter Notebook

```
In [430]: df.isna().sum()
Out[430]: type
           title
                            0
           director
                            0
           cast
           country
           date_added
           release_year
                            0
           rating
           duration
                            0
           listed_in
           week_no
                            0
           month
                            0
           year
           dtype: int64
```

In [431]: #Finally data is clean

In [432]: #Now count plot

In [433]: sns.countplot(x ="type", data = df) plt.show()



In [434]: #2.a. Find the number of movies produced in each country and pick the top 10 countries.

```
In [435]: | df[df.type=="Movie"].groupby("country")["title"].nunique().sort_values(ascending = False).head(10)
```

Out[435]: country

United States 2751 India 962 United Kingdom 532 unknown_country 440 Canada 319 France 303 182 Germany Spain 171 Japan 119 China 114 Name: title, dtype: int64

In [436]: #2.b. Find the number of Tv-Shows produced in each country and pick the top 10 countries.

```
In [437]: |df[df.type=="TV Show"].groupby("country")["title"].nunique().sort_values(ascending = False).head(10)
Out[437]: country
            United States
                                   938
            unknown_country
                                   391
            United Kingdom
                                   272
            Japan
                                   199
                                   170
            South Korea
            Canada
                                   126
            France
                                    90
            India
                                    84
                                    70
            Taiwan
            Australia
                                    66
            Name: title, dtype: int64
In [438]: |df["country"].unique()
Out[438]: array(['United States', 'South Africa', 'unknown_country', 'India',
                     'Ghana', 'Burkina Faso', 'United Kingdom', 'Germany', 'Ethiopia',
                     'Czech Republic', 'Mexico', 'Turkey', 'Australia', 'France', 'Finland', 'China', 'Canada', 'Japan', 'Nigeria', 'Spain',
                     'Belgium', 'South Korea', 'Singapore', 'Italy', 'Romania', 'Argentina', 'Venezuela', 'Hong Kong', 'Russia', '', 'Ireland',
                     'Nepal', 'New Zealand', 'Brazil', 'Greece', 'Jordan', 'Colombia',
                     'Switzerland', 'Israel', 'Taiwan', 'Bulgaria', 'Algeria', 'Poland',
                     'Saudi Arabia', 'Thailand', 'Indonesia', 'Egypt', 'Denmark',
                     'Kuwait', 'Netherlands', 'Malaysia', 'Vietnam', 'Hungary',
                     'Sweden', 'Lebanon', 'Syria', 'Philippines', 'Iceland',
                     'United Arab Emirates', 'Norway', 'Qatar', 'Mauritius', 'Austria', 'Cameroon', 'Palestine', 'Uruguay', 'United Kingdom,', 'Kenya', 'Chile', 'Luxembourg', 'Cambodia', 'Bangladesh', 'Portugal',
                     'Cayman Islands', 'Senegal', 'Serbia', 'Malta', 'Namibia',
                     'Angola', 'Peru', 'Mozambique', 'Cambodia,', 'Belarus', 'Zimbabwe',
                     'Puerto Rico', 'Pakistan', 'Cyprus', 'Guatemala', 'Iraq', 'Malawi',
                     'Paraguay', 'Croatia', 'Iran', 'West Germany', 'United States,', 'Albania', 'Georgia', 'Soviet Union', 'Morocco', 'Slovakia',
                     'Ukraine', 'Bermuda', 'Ecuador', 'Armenia', 'Mongolia', 'Bahamas', 'Sri Lanka', 'Latvia', 'Liechtenstein', 'Cuba', 'Nicaragua',
                     'Poland,', 'Slovenia', 'Dominican Republic', 'Samoa', 'Azerbaijan',
                     'Botswana', 'Vatican City', 'Jamaica', 'Kazakhstan', 'Lithuania',
                     'Afghanistan', 'Somalia', 'Sudan', 'Panama', 'Uganda',
                     'East Germany', 'Montenegro'], dtype=object)
In [439]: #3.a.Find which is the best week to release the Tv-show or the movie. Do the analysis separately for Tv-
In [440]: |df[df.type=="Movie"].groupby("week_no")["title"].nunique().sort_values(ascending = False).head(10)
Out[440]: week_no
            1
                   316
            44
                   243
            40
                   215
            9
                   207
            26
                   195
                   189
            35
            31
                   185
            13
                   174
            18
                   173
            27
                   154
            Name: title, dtype: int64
In [441]: #TV Shows
In [442]: |df[df.type=="TV Show"].groupby("week_no")["title"].nunique().sort_values(ascending = False).head(10)
Out [442]: week no
            27
                   86
            31
                   83
            13
                   76
            44
                   75
            24
                   75
            35
                   74
            5
                   73
            26
                   73
                   72
            40
            50
                   70
            Name: title, dtype: int64
```

```
In [443]: |#3.b.Find which is the best month to release the Tv-show or the movie. Do theanalysis separately for Tv-
In [444]: | #movies
In [445]: |df[df.type=="Movie"].groupby("month")["title"].nunique().sort_values(ascending = False).head(10)
Out[445]: month
          Jul
                  565
                  550
          Apr
          Dec
                  547
          Jan
                  546
          0ct
                  545
                  529
          Mar
                  519
          Aug
          Sep
                  519
          Nov
                  498
                  492
          Jun
          Name: title, dtype: int64
In [446]: #TV Shows
In [447]: |df[df.type=="TV Show"].groupby("month")["title"].nunique().sort_values(ascending = False).head(10)
Out[447]: month
                  266
          Dec
          Jul
                  262
          Sep
                  251
          Aug
                  236
          Jun
                  236
          0ct
                  215
          Apr
                  214
          Mar
                  213
                  207
          Nov
                  202
          Jan
          Name: title, dtype: int64
In [448]: | #4. Analysis of actors/directors of different types of shows/movies.
          #a. Identify the top 10 directors who have appeared in most movies or TV shows.
In [449]: |df[df.type=="Movie"].groupby("director")["title"].nunique().sort_values(ascending = False).head(10)
Out[449]: director
          Unkown_director
                                     188
          Rajiv Chilaka
                                      19
          Raúl Campos, Jan Suter
                                      18
          Suhas Kadav
                                      16
          Marcus Raboy
                                      15
          Jay Karas
                                      14
          Cathy Garcia—Molina
                                      13
          Youssef Chahine
                                      12
          Martin Scorsese
                                      12
          Jay Chapman
                                      12
          Name: title, dtype: int64
In [450]: |df[df.type=="TV Show"].groupby("director")["title"].nunique().sort_values(ascending = False).head(10)
Out[450]: director
          Unkown director
                                                                                      2446
          Alastair Fothergill
                                                                                         3
          Rob Seidenglanz
                                                                                         2
          Shin Won-ho
                                                                                         2
          Iginio Straffi
          Hsu Fu-chun
                                                                                         2
          Stan Lathan
                                                                                         2
          Ken Burns
                                                                                         2
          Pali Yahya
                                                                                         1
          Norm Hiscock, Gary Howsam, Mike Smith, John Paul Tremblay, Robb Wells
                                                                                         1
          Name: title, dtype: int64
In [451]: #4.b.Identify the top 10 actors who have appeared in most movies or TV shows.
```

```
In [452]: |df[df.type=="Movie"].groupby("cast")["title"].nunique().sort_values(ascending = False).head(10)
Out[452]: cast
          Unkown_cast
                               475
          Anupam Kher
                                42
          Shah Rukh Khan
                                35
          Naseeruddin Shah
                                32
          Om Puri
                                30
          Akshay Kumar
                                30
          Paresh Rawal
                                28
          Julie Tejwani
                                28
          Amitabh Bachchan
                                28
                                27
          Rupa Bhimani
          Name: title, dtype: int64
In [453]: |df[df.type=="TV Show"].groupby("cast")["title"].nunique().sort_values(ascending = False).head(10)
Out[453]: cast
          Unkown_cast
                                 350
          Takahiro Sakurai
                                  25
                                  19
          Yuki Kaji
          Junichi Suwabe
                                  17
          Daisuke Ono
                                  17
          Ai Kayano
                                  17
          Yuichi Nakamura
                                  16
          Jun Fukuyama
                                  15
          Yoshimasa Hosoya
                                  15
          David Attenborough
          Name: title, dtype: int64
In [488]: #5.Which genre movies are more popular or produced more
In [496]: from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
In [504]: | stopwords = set(STOPWORDS)
          wordcloud = WordCloud(width = 800, height = 400,
                           background_color ='white',
                           stopwords = stopwords,
                           min_font_size = 10).generate(" ".join(df.listed_in))
          plt.figure(figsize = (8, 8), facecolor = None)
          plt.imshow(wordcloud)
          plt.axis("off")
          plt.tight_layout(pad = 0)
          plt.show()
```

```
Adventure Comedies Independent Movies Movies Thrillers Dramas Shows Anime Shows Crime Shows Spanish Movies Action Thrillers Dramas Shows Anime Shows Crime Shows Spanish Movies Action Adventure Tv Comedies Movies International Tv Comedies Movies International Movies Independent Shows Korean Tv Shows Tv Tv Action Tv Action Tv Action Movies Romantic Movies Independent Shows Korean Tv Shows Korean T
```

In []: #from above word cloud we can interpret international movies and international tv shows are the most pop

```
In [ ]: |wordcloud = WordCloud(width = 800, height = 800,
                             background_color ='white',
                             stopwords = stopwords,
                             min_font_size = 10).generate(comment_words)
  In [ ]: |#6.Find After how many days the movie will be added to Netflix after the release ofthe movie (you can co
In [462]: |df["release_year"] = pd.to_datetime(df["release_year"], format='%Y')
In [464]: |df["diff_months"] = ((df.date_added- df.release_year)/np.timedelta64(1, 'M'))
           df["diff_months"] = df["diff_months"].astype(int)
In [465]: | df.head()
Out [465]:
                                   director
                                                 cast country date_added release_year rating duration
                type
                        title
                                                                                                      listed_in week_no month year
                        Dick
                                                       United
                                                                                     PG-
            0 Movie Johnson
                             Kirsten Johnson Unkown_cast
                                                              2021-09-25
                                                                         2020-01-01
                                                                                           90 min Documentaries
                                                                                                                        Sep 2021
                                                                                                                   38
                                                       States
                                                                                      13
                     Is Dead
                 TV
                     Blood &
                                                       South
                                                                                     TV-
                                                                                               2
                                                                                                   International
                                                              2021-09-24
                            Unkown_director
                                          Ama Qamata
                                                                         2021-01-01
                                                                                                                   38
                                                                                                                        Sep 2021
               Show
                       Water
                                                        Africa
                                                                                     MA Seasons
                                                                                                     TV Shows
                                                                                     TV-
                 TV
                     Blood &
                                                        South
                            Unkown_director
                                          Ama Qamata
                                                              2021-09-24
                                                                         2021-01-01
                                                                                                    TV Dramas
                                                                                                                   38
                                                                                                                        Sep 2021
               Show
                       Water
                                                        Africa
                                                                                     MΑ
                                                                                         Seasons
                 TV
                     Blood &
                                                                                     TV-
                                                        South
                                                                                               2
                            Unkown_director
                                          Ama Qamata
                                                              2021-09-24
                                                                         2021-01-01
                                                                                                   TV Mysteries
                                                                                                                   38
                                                                                                                        Sep 2021
               Show
                       Water
                                                        Africa
                                                                                     MA
                                                                                         Seasons
                                                                                     TV-
                                                                                                   International
                 ΤV
                     Blood &
                                                        South
                                                                                               2
                                                              2021-09-24
                                                                                                                        Sep 2021
                            Unkown_director Khosi Ngema
                                                                         2021-01-01
                                                                                                                   38
               Show
                                                                                                     TV Shows
                       Water
                                                        Africa
                                                                                     MA Seasons
In [467]: |df[df.type=="Movie"]["diff_months"].mode()
Out [467]: 0
           Name: diff_months, dtype: int64
In [471]: df[(df.type=="Movie")&(df.release_year>="2000-01-01")]["title"].nunique()
Out [471]: 5655
In [477]: df[(df.type=="Movie")&(df.release_year>="2000-01-01")]["diff_months"].nunique()
Out [477]: 249
In [481]: df[(df.type=="Movie")&(df.release_year>="2000-01-01")]["diff_months"].mode()[0]
Out[481]: 11
  In []: |#so after analyzing it seems that after 11 months from the date of release it will be ideal time to add
  In [ ]: #visualization
  In [ ]: |#word cloud
```

In []: #Genre

```
British TV
                                                                              TV TV
Sci
                                                                            ramas
                      Comedies
                                   Comedies
                                                                                Shows Crime Spanish Language
                                                                                                 Romantic TV
Shows Spanish
                            antasy<u>A</u>dventure
                                                                                               Dramas
                                                                              Comedies
      Dramas
                                                              ction
                                                   Movies
Anime Series
                                                                                                S Romantic
•Movies Music >
                           hildren
venture Dramas
                                                                   Thrillers
                                                                                    Shows Anime
                                                                                         Horror Movies Toramas Romantic B
                                                           Movies Independent
                nows
                                                 ona
    Comedies Children
```

```
In [ ]: #Country
```

```
Colombia Colombia
                                                                                                                                                                                      Thailand Thailand India United
                                                                                                                                                                                                                                                                                                                                                                                                                                                             Indonesia Indonesia
                                                                                                                                                                                                                                States India
                                                                                                                                                                                                                                                                                                                                                                                                              Arab EmiratesStates
                                                                                                                                                                                     Australia United
                                                                                                                                                                                                                                                                                                                                                                 Mexico United Braz
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              Canada Czech Republic
                                       ada
                                                                                                                                                                                                                                                                                                                                                                                                 China Hong Canada
                                                                                                                                                                                                                                         Spain United
                                                                                                                                                                                                                                                                                                                                                                                           Australia
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    States France
                                                                                                                                                                                                                                                                                                               Australia
                                                                                                                                                                                                      Turkey
France Kingdom Germ States China Germ Germ China Germ Germ China G
                                                                                                                                                                                                                       Kong China
Africa South
                                                                                                                                                                                                                                                                                                                                                               Spain
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  German
                                                                                                                                                                                                                                                                                              Spain
                                                                                                                                                                                                                                                                                                                                                                                                                                                                      Germany United
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Arge
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 Emirates United
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          Korea
Argentina
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Germany
                   Kingdom Canada
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        apan
                                                                                                                                                                                                                                                    South Africa
France Egypt
```

```
In [ ]:
```

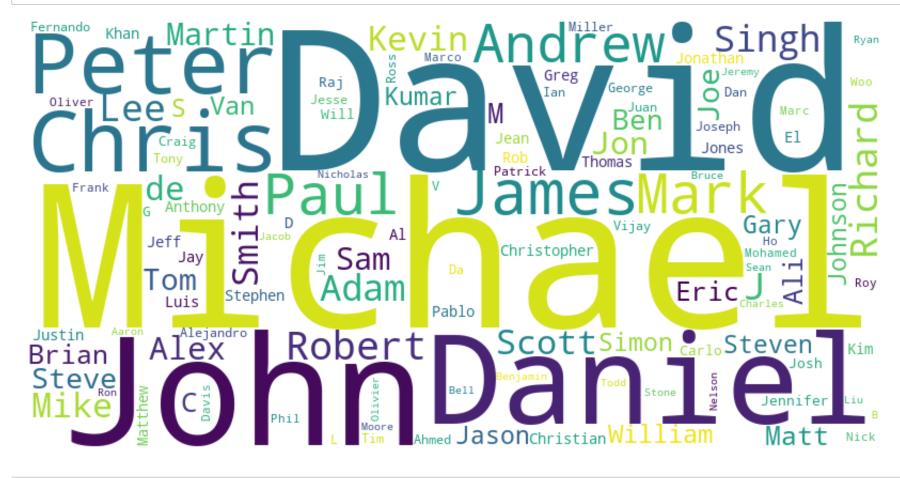
#Actors

```
Aamir Khan Colin Farrell Juan Carros Willem Dafoe
                                                                                                                                                                                                                Pierce Brosnan Yen Donnie Kristen Stewart Rajkummar Rao
                                                                                                                                                                                                       Boman Irani

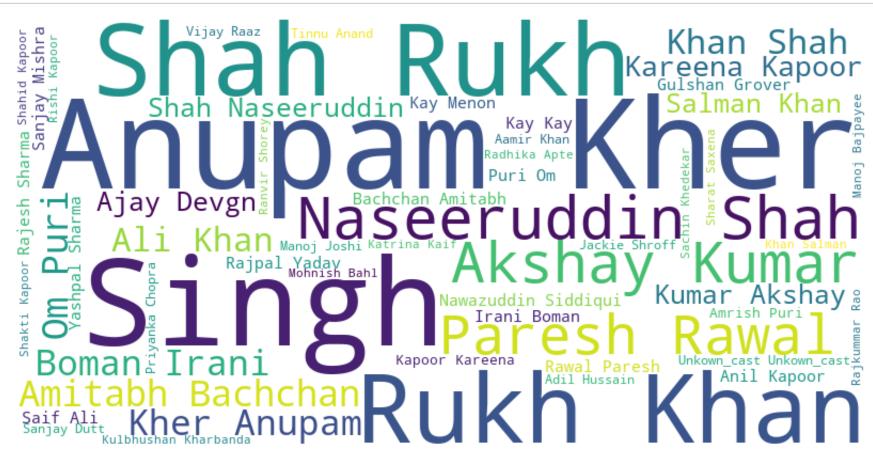
John C Gulshan Grover Kim Min

John AbrahamBrian Drummond + Nicolas Ca
         Jim Broadbent Un
                                                                                                              Brendan Gleeson
                                                                                                                                                                                                                                                                                                                                                                            ►Nicolas Cage
Nicole Kidman Vikander Alicia Sophie Kennedy Shia LaBeouf Shah Naseeruddin Shah Salman Khan Alicia Vikander de la<sub>Steve Coogan</sub> Shah Naseeruddin Saif Ali Kareena Kapoor Uma Thurman Irani Boman Nawazuddin SiddiquiBen Kingsley Tara Strong Khan Shah Cleese John Attenborough David Attenborough David Helena Bonham Raghuvir Yadav Kay Menon Mamoru Miyano Anil Kapoor Ben Whishaw Stellan Skarsgård Antonio Banderas Anton Yelche Blanchett
                                                                                                                                                                                                                                                                                                                                                                                 Nicholas Hoult
         Neeson Liam Natalie Dormer Sanjay Dutt Manoj Joshi Eddie Marsan Skarsgård Stellan David Attenborough Kaji Yuki Guy Pearce Jackie Shroff Linu Anand Carter Akshay Kumar Om Puri Bachchan Amitabh Robin Wright Lea Seydoux Carter Harvey Adil Hussain Anthony Hopkins Strong Tara Priyanka Chopra Strong Tara Priyanka C
                                                                                                                                                                                                                                                                                                                                                                                                                                     و <u>و</u>
                                                                                                                                                                                                                                                                                                                                                                         vey Keitel XX
                                                                Kristin Scott Kate Harbour Ajay Devgn Sanjay Mishra Juan Pablo
                                                   Clive Owen Amitabh Bachchan Manoj Bajpayee
                                                         Sam Worthington
                                                                                                                                                                                                                                                                       Rawal Paresh Jason Isaacs
                                                      Dormer Natalie Paresh Rawal Paresh Rawal Paresh Helen Mirren Whishaw Ben Fred Armisen Rajpal Yadav Leonardo Sbaraglia Schoenaerts Matthias
```

In []: #Director



In []: #popular actors from india



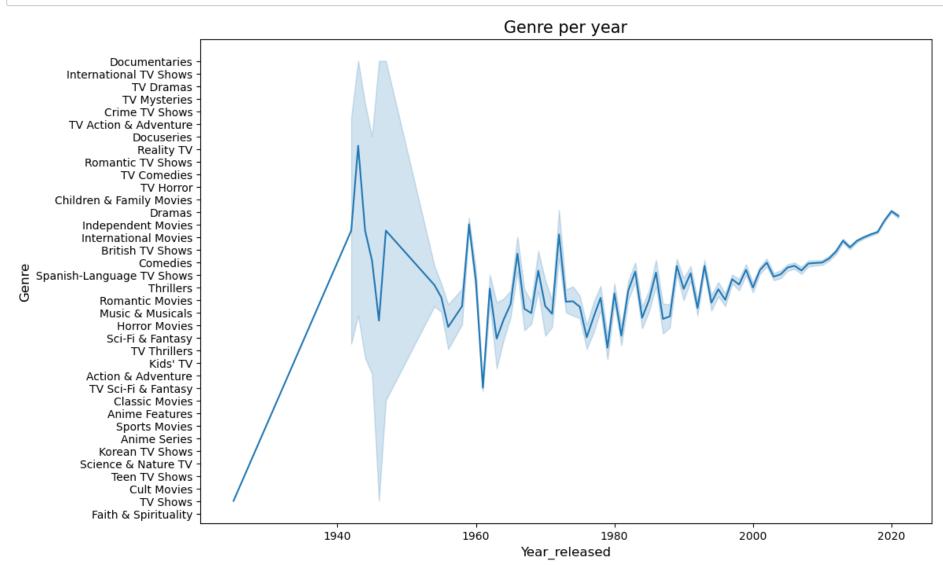
```
In []:
```

```
#popular directors from india
```

```
Mani Ratnam Bhatt Mahesh Rakeysh Omprakash Akhtar Farhan Raj Kapoor Hrishikesh Mukherjee Ghosh Ken Luthria Kulkarni Chandrakant Ratnam Mani Luthria Shah Kundan Ken Ghosh Milan Luthria Bhardwaj Vishal Pandey Neeraj Vishal Bhardwaj Vajaha Leena Jha Prakash Mehra Umesh Satish Kashyap Shardwaj Vajaha Kashyap Shardwaj Vajaha Raj Kashyap Shardwaj Vajaha Raj Kashyap Shardwaj Varma Ram Gopal Mirza Aziz Kashyap Dibakar Barjatya Sooraj Indra Kumar Mehra Prakash Kumar Indra Kashoror Raj Farah Khan Sundan Shah Mehra Rajiv Qaushiq Mukherjee Dibakar Banerjee Rohit Shetty Kukunoor Nagesh Sundan Shah Manjrekar Mahesh Banerjee Dibakar Mehra Palash Kumar Burmawalla Burmawalla Abbas Alibhai Burmawalla Rajiv Qaushiq Mukherjee Hrishikesh Shardwaj Shard
```

In []: #movie genre released per year

```
In [549]: plt.figure(figsize = (12, 8))
    sns.lineplot(x="release_year", y="listed_in", data =df)
    plt.title("Genre per year", fontsize = 15)
    plt.xlabel("Year_released", fontsize = 12)
    plt.ylabel("Genre" , fontsize = 12)
    plt.show()
```



In [576]: | df[df.type=="Movie"].groupby("release_year")["title"].nunique().sort_values(ascending=False).head(20) Out[576]: release_year 2018-01-01 767 767 2017-01-01 2016-01-01 658 2019-01-01 633 2020-01-01 517 2015-01-01 398 2021-01-01 277 2014-01-01 264 2013-01-01 225 2012-01-01 173 154 2010-01-01 2011-01-01 145 2009-01-01 118 2008-01-01 113 2006-01-01 82 2007-01-01 74 2005-01-01 67 2004-01-01 55

2002-01-01 44 Name: title, dtype: int64

2003-01-01

51

In [569]: df.head()

Out[569]:

	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	week_no	month	year
0	Movie	Dick Johnson Is Dead	Kirsten Johnson	Unkown_cast	United States	2021-09-25	2020-01-01	PG- 13	90 min	Documentaries	38	Sep	2021
1	TV Show	Blood & Water	Unkown_director	Ama Qamata	South Africa	2021-09-24	2021-01-01	TV- MA	2 Seasons	International TV Shows	38	Sep	2021
1	TV Show	Blood & Water	Unkown_director	Ama Qamata	South Africa	2021-09-24	2021-01-01	TV- MA	2 Seasons	TV Dramas	38	Sep	2021
1	TV Show	Blood & Water	Unkown_director	Ama Qamata	South Africa	2021-09-24	2021-01-01	TV- MA	2 Seasons	TV Mysteries	38	Sep	2021
1	TV Show	Blood & Water	Unkown_director	Khosi Ngema	South Africa	2021-09-24	2021-01-01	TV- MA	2 Seasons	International TV Shows	38	Sep	2021

In [577]: | dt=pd.read_csv("netflix.csv")

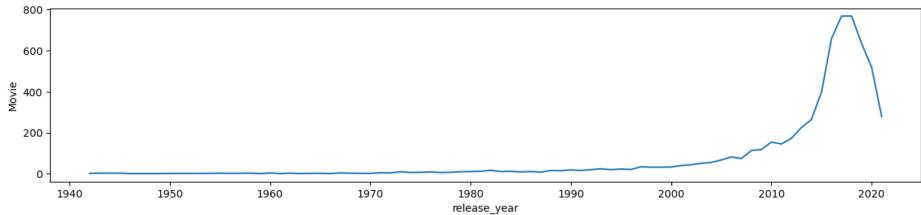
In [578]: dt.head()

Out [578]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	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	s 5	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 I

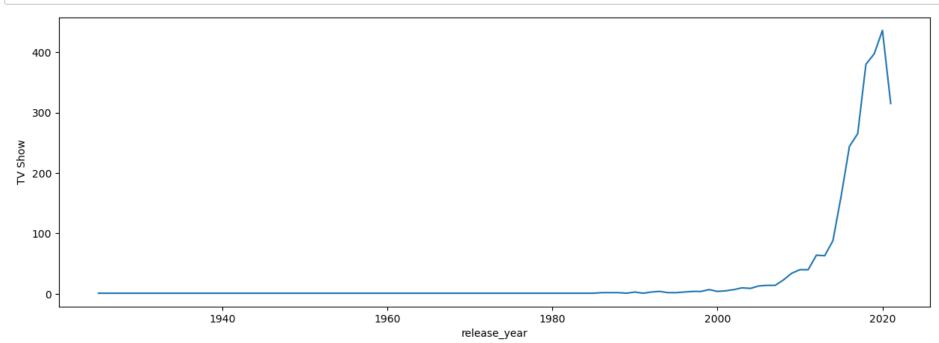
```
In [716]: |dt["director"].fillna("Unkown_director", inplace= True)
            dt["cast"].fillna('Unkown_cast', inplace= True)
            dt["country"].fillna("unknown_country",inplace= True)
            dt["duration"].fillna(df["duration"].mode()[0],inplace=True)
            dt["date_added"].fillna(df["date_added"].mode()[0],inplace=True)
            dt["rating"].fillna(df["rating"].mode()[0],inplace=True)
In [580]: | dt.head()
                                                             Qamata,
                                                                                                                              ınternationai
                                                                                      September
                                                                                                                             TV Shows, TV
                           TV
                                 Blood &
                                                               Khosi
                                                                                                               TV-
                                                                                                                         2
                                         Unkown director
                                                                         South Africa
                                                                                                       2021
                     s2
                         Show
                                   Water
                                                          Ngema, Gail
                                                                                       24, 2021
                                                                                                               MΑ
                                                                                                                   Seasons
                                                                                                                               Dramas, TV
                                                           Mabalane,
                                                                                                                                Mysteries
                                                                                                                                          Ca
                                                            Thaban...
                                                               Sami
                                                                                                                                 Crime TV
                                                                                                                                           To
                                                             Bouajila,
                                                                                                                                  Shows,
                                                               Tracy
                                                                                      September
                                                                                                               TV-
             2
                               Ganglands
                                           Julien Leclercq
                                                                                                       2021
                                                                                                                              International
                                                                     unknown_country
                                                                                       24, 2021
                                                                                                                    Season
                                                             Gotoas,
                                                                                                               MΑ
                                                                                                                             TV Shows, TV
                                                         Samuel Jouy,
                                                                                                                                    Act...
                                                                                                                                           С
                                                              Nabi...
                                 Jailbirds
                           TV
                                                                                      September
                                                                                                               TV-
                                                                                                                              Docuseries,
                                                                                                                         1
                                                                                                                                           а
             3
                                    New
                                         Unkown_director Unkown_cast unknown_country
                                                                                                       2021
                                                                                                                                Reality TV
                         Show
                                                                                       24, 2021
                                                                                                                    Season
                                  Orleans
                                                         Mayur More,
                                                                                                                                           ln
                                                                                                                              International
                                                             Jitendra
                                                                                                                                           C
                                                                                      September
                                                                                                               TV-
                                                                                                                         2
                                                                                                                               TV Shows,
                           TV
                                    Kota
                                         Unkown director
                                                             Kumar,
                                                                               India
                                                                                                       2021
                     s5
                         Show
                                  Factory
                                                                                       24, 2021
                                                                                                               MA
                                                                                                                   Seasons
                                                                                                                              Romantic TV
                                                          Ranjan Raj,
                                                                                                                             Shows, TV ...
                                                            Alam K...
In [591]:
            dt["release_year"] = pd.to_datetime(dt["release_year"], format='%Y').dt.year
In [613]:
            plt.figure(figsize = (20, 8))
            sns.countplot(x="release_year",data=dt)
            plt.title("movies/tv show released per year", fontsize = 15)
            plt.xlabel("release_year", fontsize = 12)
            plt.ylabel("movie_count" , fontsize = 12)
plt.xticks(rotation = 90, fontsize = 15)
            plt.yticks(fontsize = 12)
            plt.show()
                                                                 movies/tv show released per year
               1200
               1000
                800
             movie_count
               600
                400
                200
                                                                           release_year
  In [ ]: #movie released per year
In [715]: | dt_n = dt.groupby("release_year")["type"].value_counts().unstack(level=-1)
```

```
In [714]: plt.figure(figsize = (15, 3))
    sns.lineplot(x="release_year", y="Movie", data=dt_n)
    plt.show()
```



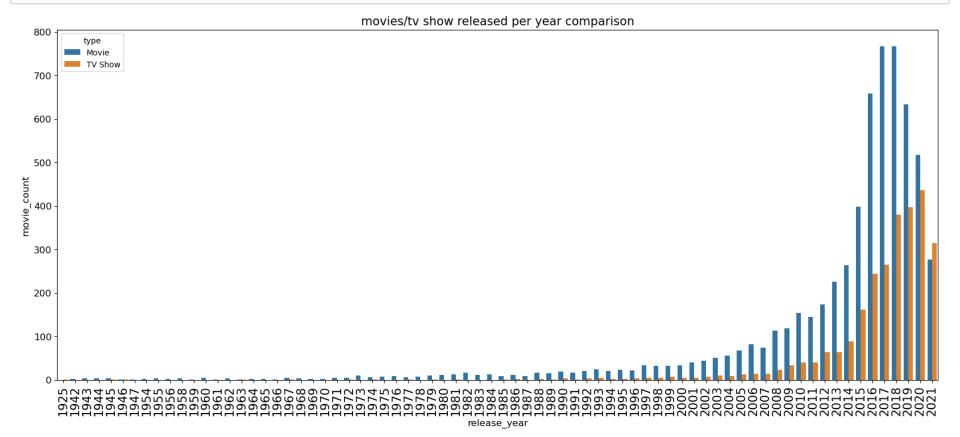
In []: #tv shows released per year

```
In [713]: plt.figure(figsize = (15, 5))
sns.lineplot(x="release_year", y="TV Show", data=dt_n)
plt.show()
```



In []: #comparison of tv shows/movies

```
In [616]: plt.figure(figsize = (20, 8))
    sns.countplot(x="release_year", hue="type", data=dt)
    plt.title("movies/tv show released per year comparison", fontsize = 15)
    plt.xlabel("release_year", fontsize = 12)
    plt.ylabel("movie_count", fontsize = 12)
    plt.xticks(rotation = 90, fontsize = 15)
    plt.yticks(fontsize = 12)
    plt.show()
```

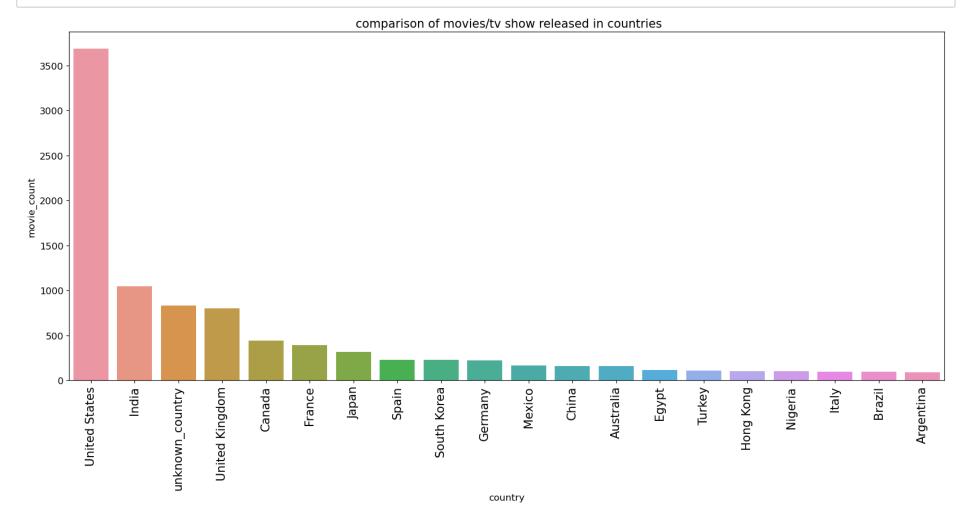


```
In []: #in 2020 highest tv shows has been released #in 2018&2017 highest movie has been released
```

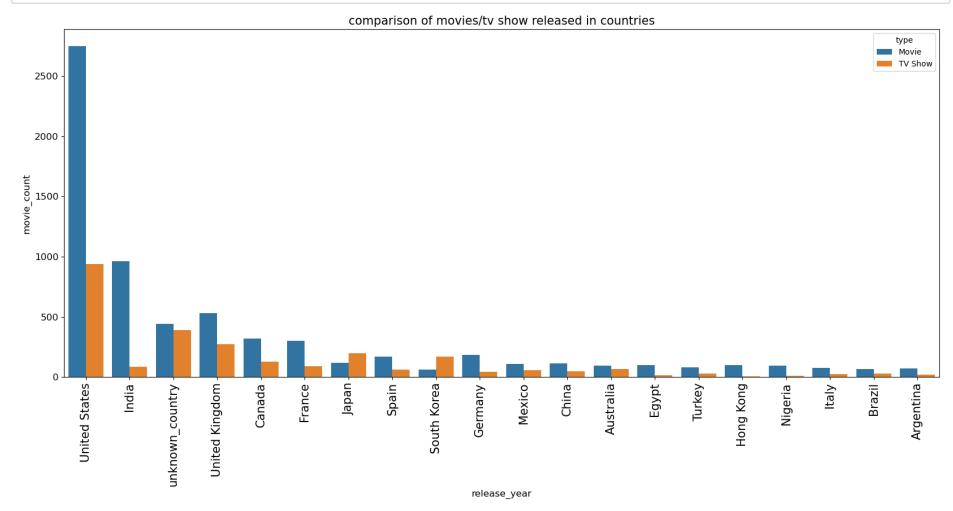
```
In [712]: dt["country"] = dt["country"].apply(lambda x: str(x).split(', '))
dt = dt.explode("country")
```

In []: #tv shows&movies comparison for countries

```
In [710]: plt.figure(figsize = (20, 8))
    sns.countplot(x="country",data=dt, order =dt["country"].value_counts().index[:20])
    plt.title(" comparison of movies/tv show released in countries",fontsize = 15)
    plt.xlabel("country", fontsize = 12)
    plt.ylabel("movie_count", fontsize = 12)
    plt.xticks(rotation = 90, fontsize = 15)
    plt.yticks(fontsize = 12)
    plt.show()
```



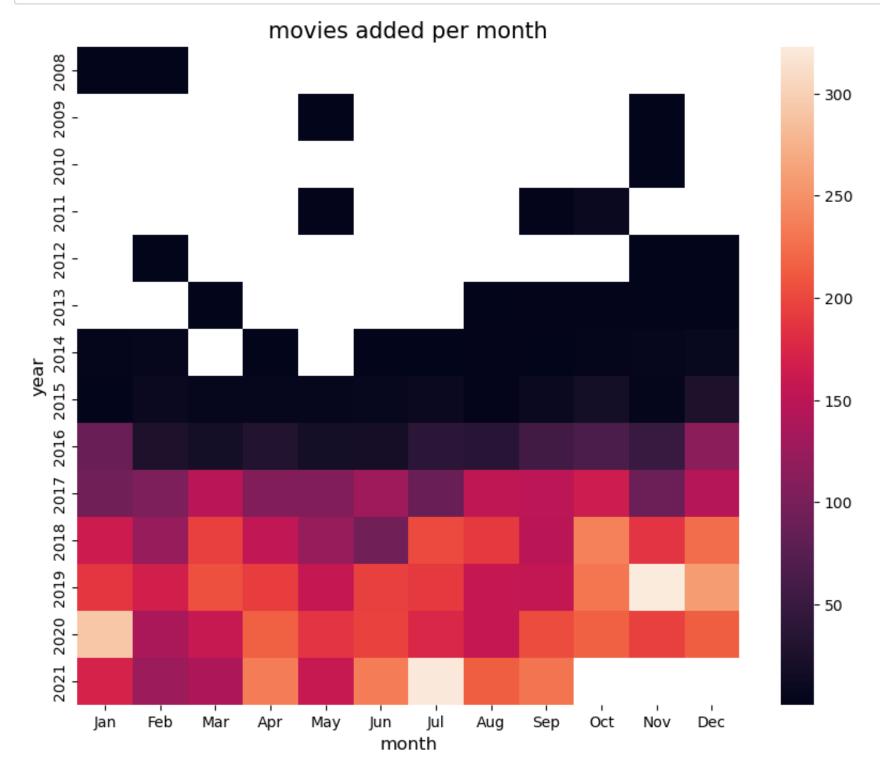
```
In [709]: plt.figure(figsize = (20, 8))
    sns.countplot(x="country", hue="type", data=dt, order =dt["country"].value_counts().index[:20])
    plt.title("comparison of movies/tv show released in countries", fontsize = 15)
    plt.xlabel("release_year", fontsize = 12)
    plt.ylabel("movie_count", fontsize = 12)
    plt.xticks(rotation = 90, fontsize = 15)
    plt.yticks(fontsize = 12)
    plt.show()
```



In []: #from the above plot we can infer that united states have highest number of movies&tv shows releases.

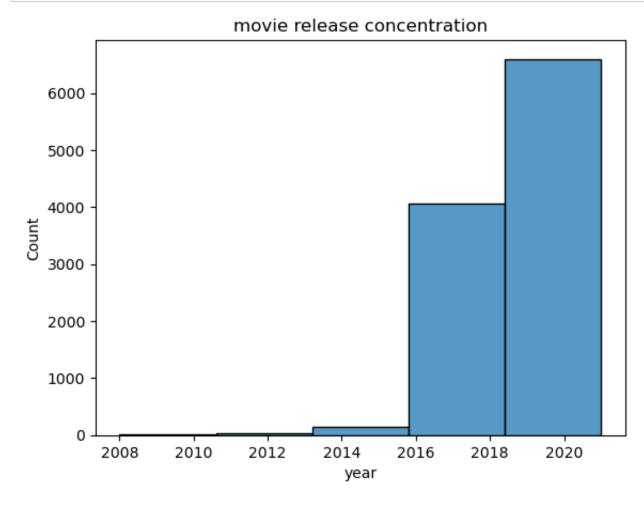
In []: #comparison of movie release in each month per year using heatmap

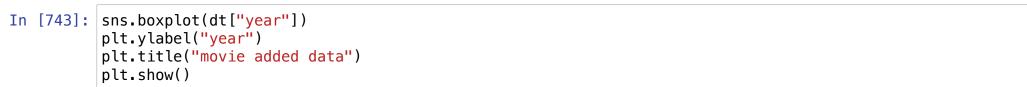
```
In [704]: plt.figure(figsize = (10, 8))
    sns.heatmap(dat)
    plt.title("movies added per month", fontsize = 15)
    plt.xlabel("month", fontsize = 12)
    plt.ylabel("year" , fontsize = 12)
    plt.show()
```

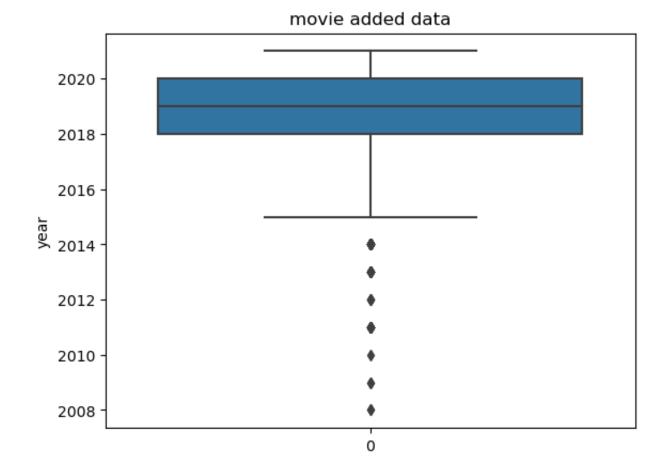


In []: #based on 2021 data we can say april, june&july is the most favourable month for movie release

```
In [742]: sns.histplot(dt["year"],bins=5)
plt.title("movie release concentration")
plt.show()
```







In []: #from above to plot we can infer that maximum movies were added between 2018 to 2020.

In []: