

Netflix - Data Exploration & Visualization

Business Problem -

- 1) which type of shows/movies to produce
- 2) how they can grow the business in different countries

```
In [143]:  
  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
df = pd.read_csv("netflix.csv") # Loading the Data set.  
df.head()
```

	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 Mablane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

Asumptions -

**As Viewers data is not provided so it we assumed that Netflix releases content on the basis of their own analysis.

**We are also assumig that the amount of content available is directly proportional to market share.

Research Questions-

1. What is more popular Shows / movies?
2. List of countries where market share for movies is higher.
3. List of countries where market share for movies is lower.
4. List of countries where market share for Shows is higher.
5. High in demand genres in last 5 years for movies.
6. High in demand genres in last 5 years for shows.
7. Least in demand genres in last 5 years for movies.
8. Least in demand genres in last 5 years for shows.
9. High demand genres from top 2 countries
10. based on the demand how to grow the market - checking shows seasons

Now, Lets start by cleaning the Data, followed by Data Pre-Processing.

```
In [144]:  
  
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
```

```
In [145]:  
  
df.shape  
  
(8807, 12)
```

Step 1: Handling NaN or missing values.

Imputation

Date_Added => Release year

country => Cast / Director

Cast => Country

Director => listed_in/country

drop_na

rating

duration

country & Director & Cast

Drop NaN Values

```
In [146]:  
  
df1 = df.dropna(subset=['rating']) # Drop nan values for rating.  
df1 = df1.dropna(subset=["duration"]) # Drop nan values for duration.  
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 8800 entries, 0 to 8806  
Data columns (total 12 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   show_id         8800 non-null   object  
1   type            8800 non-null   object  
2   title           8800 non-null   object  
3   director        6169 non-null   object  
4   cast            7975 non-null   object  
5   country         7970 non-null   object  
6   date_added      8790 non-null   object  
7   release_year    8800 non-null   int64  
8   rating          8800 non-null   object  
9   duration        8800 non-null   object  
10  listed_in       8800 non-null   object  
11  description      8800 non-null   object  
dtypes: int64(1), object(11)  
memory usage: 893.8+ KB
```

```
In [147]:

null_df = df1[['country','cast','director']].isna()
null_df = null_df[ (null_df['country']==True) & (null_df['cast']==True) & (null_df['director']==True) ]
index_to_drop = null_df.index

df1 = df1.drop(index=index_to_drop) # Drop nan values where country, cast & director all 3 are missing.
df1
```

	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 fat
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 C
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 protec family fro powerful .lor...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city c coaching centers known to train l...
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	The arriv a charism young pri brin...
...
8801	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Recoveri alcoholic Talal wak up inside s...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonis crime reporter a...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking t survive in world tak over by z
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civili life, a forr superheri
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrapp poor boy worms hi way into i ty...

8704 rows × 12 columns

Imputation

```
In [148]:
df1["date_added"] = df1["date_added"].fillna(df1["release_year"]) # Imputing date_added with release year
df1["date_added"].isna().sum()
```

0

```
In [149]:
df1.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 8704 entries, 0 to 8806
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   show_id         8704 non-null   object
1   type            8704 non-null   object
2   title           8704 non-null   object
3   director        6169 non-null   object
4   cast            7975 non-null   object
5   country         7970 non-null   object
6   date_added      8704 non-null   object
7   release_year    8704 non-null   int64
8   rating          8704 non-null   object
9   duration        8704 non-null   object
10  listed_in       8704 non-null   object
11  description     8704 non-null   object
dtypes: int64(1), object(11)
memory usage: 884.0+ KB
```

Imputation of directors

```
In [150]:
director_fill = df1.groupby(['type', "listed_in", "country"])[ 'director' ].apply(lambda x: x.mode())
director_fill = director_fill.reset_index()
director_fill
```

	type	listed_in	country	level_3	director
0	Movie	Action & Adventure	Australia, United States	0	Simon Wincer
1	Movie	Action & Adventure	Bulgaria, United States	0	Todor Chapkanov
2	Movie	Action & Adventure	Canada	0	Adam Alleca
3	Movie	Action & Adventure	Canada	1	Allan Ungar
4	Movie	Action & Adventure	Canada	2	Brett Donowho
...
3195	TV Show	TV Shows	Japan	0	Kobun Shizuno, Hiroyuki Seshita
3196	TV Show	TV Shows	Saudi Arabia	0	Ali Kalthami, Meshal Aljaser, Faris Godus, Moh...
3197	TV Show	TV Shows	United States	0	Joe Berlinger, Bruce Sinofsky
3198	TV Show	TV Shows	United States	1	Oscar Micheaux, Spencer Williams, Richard E. N...
3199	TV Show	TV Shows	United States	2	Steven Bognar, Julia Reichert

3200 rows × 5 columns

```
In [151]:  
  
# Imputing 1st level = ["type", "listed_in", "country" ]  
  
df1 = df1.merge(director_fill, on = ["type", "listed_in", "country" ] , how = 'left')  
df1["director"] = np.where(df1["director_x"].isna(), df1["director_y"], df1["director_x"])  
df1 = df1[["show_id", "type", "title", "director", "cast", "country", "date_added", "release_year", "rating", "duration", "listed_in", "description"]]  
df1 = df1.drop_duplicates( subset='show_id' )  
df1
```

	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 fi nears th end of h life, filmi
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 party, a 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 prote family fr powerfu lor...
3	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city coaching centers known t train l...
4	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	The arri a charis young p prin...
...
29393	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Recover alcoholic Talal wa up insidi s...
29394	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A politic: cartooni crime reporter a...
29395	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking survive i world ta over by
29397	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civ life, a fo superhe
29398	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrap; poor boy worms i way into ty...

8704 rows × 12 columns

```
In [152]:  
  
# Imputing 2nd level = ["type","listed_in"]  
  
if2 = df1.merge(director_fill[['type','listed_in','director']], on = ["type","listed_in"] , how = 'left')  
if2["director"] = np.where(df2["director_x"].isna(), df2["director_y"], df2["director_x"])  
if2 = df2[["show_id","type","title","director","cast","country","date_added","release_year","rating","durati  
if2 = df2.drop_duplicates( subset='show_id' )  
if2
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	desc
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As he nears end c life, fi
48	s2	TV Show	Blood & Water	Michel Tikhomiroff	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 cross paths party, Town
52	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 pn family powe lor...
54	s5	TV Show	Kota Factory	BB Sasore	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 c coact cente know train l
60	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	The a a cha young brin...
...
357343	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Reco alcoh Talal up ins s...
357457	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A poli cartor crime repor a...
357458	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looki surviv world over l
357462	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragt from i life, a super
357542	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scr poor l worm way ii ty...

8704 rows × 12 columns

```
In [153]:  
  
# Imputing 3rd level = ["type"]  
  
df3 = df2.merge(director_fill[["type", "director"]], on = ["type"], how = 'left')  
df3["director"] = np.where(df3["director_x"].isna(), df3["director_y"], df3["director_x"])  
df3 = df3[["show_id", "type", "title", "director", "cast", "country", "date_added", "release_year", "rating", "duration", "listed_in", "description"]]  
df3 = df3.drop_duplicates(subset='show_id')  
df3
```

	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 ne: en: life
3033	s2	TV Show	Blood & Water	Michel Tikhomiroff	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	Aft cro pal par Toi
3200	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 fan pov lor.
3367	s5	TV Show	Kota Factory	BB Sasore	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 i co: cer kn: trai
3534	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	Thi a c you brii
...
18946797	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Re alc Tal up s...
18949830	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A f car crir rep a...
18952863	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Lo: sur wo ove
18955896	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dre: froi life suq
18958929	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A s pot wo wa ty..

8704 rows × 12 columns


```
In [154]:  
  
# Reseting df1 & df2 with latest df3  
df1=df3  
df2=df3  
  
df3.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 8704 entries, 0 to 18958929  
Data columns (total 12 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   show_id         8704 non-null   object  
1   type            8704 non-null   object  
2   title           8704 non-null   object  
3   director        8704 non-null   object  
4   cast            7975 non-null   object  
5   country         7970 non-null   object  
6   date_added      8704 non-null   object  
7   release_year    8704 non-null   int64  
8   rating          8704 non-null   object  
9   duration        8704 non-null   object  
10  listed_in       8704 non-null   object  
11  description     8704 non-null   object  
dtypes: int64(1), object(11)  
memory usage: 884.0+ KB
```

Imputation of Cast

```
In [155]:  
  
cast_fill = df1.groupby(['type','listed_in','country'])['cast'].apply(lambda x: x.mode())  
cast_fill = cast_fill.reset_index()  
cast_fill
```

	type	listed_in	country	level_3	cast
0	Movie	Action & Adventure	Australia, United States	0	Tom Selleck, Laura San Giacomo, Alan Rickman, ...
1	Movie	Action & Adventure	Bulgaria, United States	0	Scott Adkins, Teodora Duhovnikova, Alon Aboutb...
2	Movie	Action & Adventure	Canada	0	Cole Hauser, Bruce Willis, Shawn Ashmore, Asht...
3	Movie	Action & Adventure	Canada	1	Dominic Purcell, Cody Hackman, Stephen Lang, T...
4	Movie	Action & Adventure	Canada	2	Jason Momoa, Garret Dillahunt, Jill Wagner, St...
...
6470	TV Show	TV Shows	India	0	Rahman, Havish, Regina Cassandra, Nandita Swet...
6471	TV Show	TV Shows	India	1	Suriya, Vijay Sethupathi, Revathy, Prakash Raj...
6472	TV Show	TV Shows	Japan	0	Mamoru Miyano, Takahiro Sakurai, Kana Hanazawa...
6473	TV Show	TV Shows	Saudi Arabia	0	Moayad Alnefaie, Mohammed Alhamdan, Citi Bet B...
6474	TV Show	TV Shows	United States	0	James Hetfield, Lars Ulrich, Kirk Hammett, Rob...

6475 rows × 5 columns

```
In [156]:  
  
# Imputing 1st level = ["type","listed_in","country" ]  
  
df1 = df1.merge(cast_fill, on = ["type","listed_in","country" ] , how = 'left')  
df1["cast"] = np.where(df1["cast_x"].isna(), df1["cast_y"], df1["cast_x"])  
df1 = df1[["show_id","type","title","director","cast","country","date_added","release_year","rating","duration"]]  
df1 = df1.drop_duplicates( subset='show_id' )  
df1
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	desc
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Craig Sechler	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As he nears end c life, fi
1	s2	TV Show	Blood & Water	Michel Tikhomiroff	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 cross paths party, Town
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 pn family powe lor...
3	s5	TV Show	Kota Factory	BB Sasore	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 c coact cente know train l
10	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	The a a cha young brin...
...
152539	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Reco alcoh Talal up ins s...
152540	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A poli cartor crime repor a...
152541	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looki surviv world over l
152551	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragt from i life, a super
152641	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scr poor l worm way ii ty...

8704 rows × 12 columns

```
In [157]:  
  
# Imputing 2nd level = ["type","listed_in"]  
  
df2 = df1.merge(cast_fill[["type","listed_in","cast"]], on = ["type","listed_in"] , how = 'left')  
df2["cast"] = np.where(df2["cast_x"].isna(), df2["cast_y"], df2["cast_x"])  
df2 = df2[["show_id","type","title","director","cast","country","date_added","release_year","rating","duration"]]  
df2 = df2.drop_duplicates( subset='show_id' )  
df2
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	desc
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Craig Sechler	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As he nears end c life, fi
23	s2	TV Show	Blood & Water	Michel Tikhomiroff	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 cross paths party, Town
47	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 pn family powe lor...
60	s5	TV Show	Kota Factory	BB Sasore	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 c coact cente know train l
126	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	The a a cha young brin...
...
697210	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Reco alcoh Talal up ins s...
697334	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A poli cartor crime repor a...
697335	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looki surviv world over l
697347	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragt from i life, a super
697527	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scr poor l worm way ii ty...

8704 rows × 12 columns

```
In [158]:  
  
# Imputing 3rd level = ["type"]  
  
df3 = df2.merge(cast_fill[["type","cast"]], on = ["type"] , how = 'left')  
df3["cast"] = np.where(df3["cast_x"].isna(), df3["cast_y"], df3["cast_x"])  
df3 = df3[["show_id","type","title","director","cast","country","date_added","release_year","rating","duration","listed_in","description"]]  
df3 = df3.drop_duplicates(subset='show_id')  
df3
```

	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	Craig Sechler	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As ne: en: c life
4716	s2	TV Show	Blood & Water	Michel Tikhomiroff	Ama Qamata, Khosi Ngema, Gail Mablane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	Aft cro pal pai Toi
6475	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 fan poi lor.
8234	s5	TV Show	Kota Factory	BB Sasore	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 i coi cer kn: tra
9993	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	Thi a c yoi bri
...
33351069	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Re alc Tal up s...
33355785	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A f car crir rep a...
33360501	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Lo: sur wo ove
33365217	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dr: froi life sur
33369933	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A s poi wo wa ty..

8704 rows × 12 columns

```
In [159]:  
  
# Reseting df1 & df2 with latest df3  
df1=df3  
df2=df3  
  
df3.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 8704 entries, 0 to 33369933  
Data columns (total 12 columns):  
#   Column                Non-Null Count  Dtype  
---  ----  
0   show_id                8704 non-null   object  
1   type                   8704 non-null   object  
2   title                  8704 non-null   object  
3   director               8704 non-null   object  
4   cast                   8704 non-null   object  
5   country                7970 non-null   object  
6   date_added             8704 non-null   object  
7   release_year           8704 non-null   int64  
8   rating                 8704 non-null   object  
9   duration               8704 non-null   object  
10  listed_in              8704 non-null   object  
11  description             8704 non-null   object  
dtypes: int64(1), object(11)  
memory usage: 884.0+ KB
```

Imputation of Country

```
In [160]:  
  
country_fill = df1.groupby(["cast", "director"])['country'].apply(lambda x: x.mode())  
country_fill = country_fill.reset_index()  
country_fill
```

		cast	director	level_2	country
0	'Najite Dede, Jude Chukwuka, Taiwo Arimoro, Od...	Jay Oliva		0	Nigeria
1	4Minute, B1A4, BtoB, ELSIE, EXID, EXO, Got7, I...	Muh Chen		0	South Korea
2	50 Cent, Ryan Phillippe, Bruce Willis, Rory Ma...	Mike Gunther		0	United States
3	A.J. LoCascio, Sendhil Ramamurthy, Fred Tatas...	Michael D. Black		0	United States
4	A.R. Rahman	Rudradeep Bhattacharjee		0	India
...
7648	İbrahim Büyükkak, Zeynep Koçak, Gupse Özay, Cen...	Bedran Güzel		0	Turkey
7649	İbrahim Çelikkol, Belçim Bilgin, Alican Yüceso...	Ehtesham Uddin		0	Turkey
7650	Şahin Irmak, İrem Sak, Gonca Vuslateri, Emre K...	Hakan Algül		0	Turkey
7651	Şükrü Özyıldız, Aslı Enver, Şenay Gürlər, Başa...	Neslihan Yesilyurt		0	Turkey
7652	Şopê Dirîsû, Wunmi Mosaku, Matt Smith, Malaik...	Remi Weekes		0	United Kingdom

7653 rows x 4 columns

```
In [161]:  
  
# Imputing 1st level = ["cast","director"]  
  
df1 = df1.merge(country_fill, on = ["cast","director"] , how = 'left')  
df1["country"] = np.where(df1["country_x"].isna(), df1["country_y"], df1["country_x"])  
df1 = df1[["show_id","type","title","director","cast","country","date_added","release_year","rating","duration"]]  
df1 = df1.drop_duplicates( subset='show_id' )  
df1
```

	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	Craig Sechler	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her f nears th end of h life, film
1	s2	TV Show	Blood & Water	Michel Tikhomiroff	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 party, a 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 prote family fr powerfu lor...
3	s5	TV Show	Kota Factory	BB Sasore	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 coachin centers known t train l...
4	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	The arri a charis young p brin...
...
8823	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Recove alcoholi Talal wa up insid s...
8824	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A politic cartooni crime reporter a...
8825	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking survive world ta over by
8826	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragger from civ life, a fo superhe
8827	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrap poor bo worms l way intc ty...

8704 rows × 12 columns

```
In [162]:  
  
# Imputing 2nd level = ["cast"]  
  
df2 = df1.merge(country_fill[["country","cast"]], on = ["cast"] , how = 'left')  
df2["country"] = np.where(df2["country_x"].isna(), df2["country_y"], df2["country_x"])  
df2 = df2[["show_id","type","title","director","cast","country","date_added","release_year","rating","duration"]]  
df2 = df2.drop_duplicates( subset='show_id' )  
df2
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descr
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Craig Sechler	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her nears l end of life, filr
125	s2	TV Show	Blood & Water	Michel Tikhomiroff	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 crossir paths i party, i Town t
126	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 prof family powerf lor...
127	s5	TV Show	Kota Factory	BB Sasore	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 cit coachi center: known train l..
128	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	NaN	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	The ar a chari young brin...
...
33174	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Recov. alcoho Talal w up insi s...
33175	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A politi cartooi crime reporte a...
33176	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Lookin survive world t over b;
33177	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragge from ci life, a f superh
33178	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scraj poor b worms way in ty...

8704 rows × 12 columns

```
In [163]:
# Imputing 3rd level = Mode of country

con_mode = df['country'].mode()[0]
df2["country"] = df2['country'].fillna(con_mode)
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8704 entries, 0 to 33178
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   show_id         8704 non-null   object
1   type            8704 non-null   object
2   title           8704 non-null   object
3   director        8704 non-null   object
4   cast            8704 non-null   object
5   country         8704 non-null   object
6   date_added      8704 non-null   object
7   release_year    8704 non-null   int64
8   rating          8704 non-null   object
9   duration        8704 non-null   object
10  listed_in       8704 non-null   object
11  description      8704 non-null   object
dtypes: int64(1), object(11)
memory usage: 884.0+ KB
```

```
In [164]:
# Resetting df1 & df3 with latest df2
df1=df2
df3=df2

df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8704 entries, 0 to 33178
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   show_id         8704 non-null   object
1   type            8704 non-null   object
2   title           8704 non-null   object
3   director        8704 non-null   object
4   cast            8704 non-null   object
5   country         8704 non-null   object
6   date_added      8704 non-null   object
7   release_year    8704 non-null   int64
8   rating          8704 non-null   object
9   duration        8704 non-null   object
10  listed_in       8704 non-null   object
11  description      8704 non-null   object
dtypes: int64(1), object(11)
memory usage: 884.0+ KB
```


Imputation and handling nan values is done.

```
In [165]:  
  
df2 = df2.reset_index(drop=True)  
df1=df2  
df1.head()
```

	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	Craig Sechler	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	Michel Tikhomiroff	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...	United States	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s5	TV Show	Kota Factory	BB Sasore	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...
4	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	United States	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries	The arrival of a charismatic young priest brin...

Step 2: Unnesting => cast , director , country and genre(listed_in)

```
In [166]:
# unnesting Genre (listed_in)

unnest_coloumn = df1["listed_in"].str.split(",", expand=True).add_prefix('listed_in') # expand = True creat

df2 = df1.merge(unnest_coloumn, left_index=True, right_index=True)

df2.drop("listed_in", axis=1, inplace=True)

df3=pd.melt(df2, id_vars=['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added', 'release
    ,value_name = 'genre']).drop("variable", axis=1) # id_vars= [list of those coloumns which we don
    # variable name ka ek column hamasa banta hai ji

df3.dropna(subset=['genre'], inplace=True)

df1 = df3
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 19130 entries, 0 to 26111
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   show_id         19130 non-null  object
 1   type            19130 non-null  object
 2   title           19130 non-null  object
 3   director        19130 non-null  object
 4   cast            19130 non-null  object
 5   country         19130 non-null  object
 6   date_added      19130 non-null  object
 7   release_year    19130 non-null  int64
 8   rating          19130 non-null  object
 9   duration        19130 non-null  object
10   description     19130 non-null  object
11   genre           19130 non-null  object
dtypes: int64(1), object(11)
memory usage: 1.9+ MB
```

```
In [167]:
# unnesting cast

unnest_coloumn = df1["cast"].str.split(", ", expand=True).add_prefix('cast') # expand = True creates coloumn

df2 = df1.merge(unnest_coloumn, left_index=True, right_index=True)

df2.drop("cast", axis=1, inplace=True)

df3=pd.melt(df2, id_vars=['show_id', 'type', 'title', 'director', 'genre', 'country', 'date_added', 'releas
            , value_name = 'cast']).drop("variable", axis=1) # id_vars= [list of those coloumns which we dont
                                                            # variable name ka ek column hamasa banta hai ji

df3.dropna(subset=['cast'], inplace=True)

df1 = df3
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 151288 entries, 0 to 954448
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ---
 0   show_id         151288 non-null object
 1   type            151288 non-null object
 2   title           151288 non-null object
 3   director        151288 non-null object
 4   genre           151288 non-null object
 5   country         151288 non-null object
 6   date_added      151288 non-null object
 7   release_year    151288 non-null int64
 8   rating          151288 non-null object
 9   duration        151288 non-null object
10   description     151288 non-null object
11   cast            151288 non-null object
dtypes: int64(1), object(11)
memory usage: 15.0+ MB
```

```
In [168]:
# unnesting director

unnest_coloumn = df1["director"].str.split(",",expand=True).add_prefix('director') # expand = True creates

df2 = df1.merge(unnest_coloumn, left_index=True, right_index=True)

df2.drop("director",axis=1,inplace=True)

df3=pd.melt(df2, id_vars=['show_id', 'type', 'title', 'genre', 'cast', 'country', 'date_added','release_ye
            ,value_name = 'director']).drop("variable",axis=1) # id_vars= [list of those coloumns which we
                                                                # variable name ka ek column hamasa banta hai ji

df3.dropna(subset=['director'],inplace=True)

df1 = df3
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 168211 entries, 0 to 1962014
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   show_id         168211 non-null object
 1   type            168211 non-null object
 2   title           168211 non-null object
 3   genre           168211 non-null object
 4   cast            168211 non-null object
 5   country         168211 non-null object
 6   date_added      168211 non-null object
 7   release_year    168211 non-null int64
 8   rating          168211 non-null object
 9   duration        168211 non-null object
10   description     168211 non-null object
11   director        168211 non-null object
dtypes: int64(1), object(11)
memory usage: 16.7+ MB
```

In [169]:

```

# unnesting country

unnest_coloumn = df1["country"].str.split(",",expand=True).add_prefix('country') # expand = True creates c

df2 = df1.merge(unnest_coloumn, left_index=True, right_index=True)

df2.drop("country",axis=1,inplace=True)

df3=pd.melt(df2, id_vars=['show_id', 'type', 'title', 'director', 'cast', 'genre', 'date_added','release_y
            ,value_name = 'country']).drop("variable",axis=1) # id_vars= [list of those coloumns which we d
                                                                # variable name ka ek column hamasa banta hai ji

df3.dropna(subset=['country'],inplace=True)

df1 = df3
df1.info()

```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 210281 entries, 0 to 1863790
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ---
 0   show_id         210281 non-null object
 1   type            210281 non-null object
 2   title           210281 non-null object
 3   director        210281 non-null object
 4   cast            210281 non-null object
 5   genre           210281 non-null object
 6   date_added      210281 non-null object
 7   release_year    210281 non-null int64
 8   rating          210281 non-null object
 9   duration        210281 non-null object
10   description     210281 non-null object
11   country         210281 non-null object
dtypes: int64(1), object(11)
memory usage: 20.9+ MB

```

```
In [170]:  
df1.head(10)
```

	show_id	type	title	director	cast	genre	date_added	release_year	rating	duration	description	count
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Craig Sechler	Documentaries	September 25, 2021	2020	PG-13	90 min	As her father nears the end of his life, filmm...	United States
1	s2	TV Show	Blood & Water	Michel Tikhomiroff	Ama Qamata	International TV Shows	September 24, 2021	2021	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...	South Africa
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila	Crime TV Shows	September 24, 2021	2021	TV-MA	1 Season	To protect his family from a powerful drug lor...	United States
3	s5	TV Show	Kota Factory	BB Sasore	Mayur More	International TV Shows	September 24, 2021	2021	TV-MA	2 Seasons	In a city of coaching centers known to train l...	India
4	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel	TV Dramas	September 24, 2021	2021	TV-MA	1 Season	The arrival of a charismatic young priest brin...	United States
5	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	Children & Family Movies	September 24, 2021	2021	PG	91 min	Equestria's divided. But a bright-eyed hero be...	United States
6	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	Dramas	September 24, 2021	1993	TV-MA	125 min	On a photo shoot in Ghana, an American model s...	United States
7	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc	British TV Shows	September 24, 2021	2021	TV-14	9 Seasons	A talented batch of amateur bakers face off in...	United Kingdom
8	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy	Comedies	September 24, 2021	2021	PG-13	104 min	A woman adjusting to life after a loss contend...	United States
9	s12	TV Show	Bangkok Breaking	Kongkiat Komesiri	Sukollawat Kanarot	Crime TV Shows	September 23, 2021	2021	TV-MA	1 Season	Struggling to earn a living in Bangkok, a man ...	United States

Step 3: Data Pre-Processing

```
In [171]:  
df1['country'].value_counts().head(10)
```

```
United States    62885  
India            23249  
United Kingdom  10138  
United States    9819  
Japan            7378  
South Korea      5766  
France           4856  
Canada           4567  
Spain            4301  
France           4104  
Name: country, dtype: int64
```

There are some dublicate records with spaces in front
Need to pre-process and clean these data as well

```
In [172]:
df1["cast"] = df1["cast"].str.strip()
df1['genre'] = df1["genre"].str.strip()
df1['director'] = df1["director"].str.strip()
df1['country'] = df1["country"].str.strip()

df1.drop_duplicates(keep="first", inplace=True)
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 210226 entries, 0 to 1863790
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   show_id         210226 non-null object
1   type            210226 non-null object
2   title           210226 non-null object
3   director        210226 non-null object
4   cast            210226 non-null object
5   genre           210226 non-null object
6   date_added      210226 non-null object
7   release_year    210226 non-null int64
8   rating          210226 non-null object
9   duration        210226 non-null object
10  description      210226 non-null object
11  country          210226 non-null object
dtypes: int64(1), object(11)
memory usage: 20.9+ MB
```

```
In [173]:
df1['country'].value_counts().head(10)
```

```
United States    72679
India            23970
United Kingdom   13466
France           8958
Japan            8784
Canada           8157
South Korea      6219
Spain            5398
Germany          4488
Mexico           4193
Name: country, dtype: int64
```

Duplicate records with Leading spaces are cleaned

Handeling Duration Feature

```
In [174]:
# Seperating movies & TV Shows

movies_df = df1[df1['type'] == 'Movie']

shows_df = df1[df1['type'] == 'TV Show']
```

```
In [175]:  
  
# Extracting minutes  
  
movies_df['duration'] = movies_df['duration'].str.split().apply( lambda x: x[0] )  
movies_df
```

	show_id	type	title	director	cast	genre	date_added	release_year	rating	duration	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Craig Sechler	Documentaries	September 25, 2021	2020	PG-13	90	As her father nears the end of his life, filmm...
5	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	Children & Family Movies	September 24, 2021	2021	PG	91	Equestria's divided. But a bright-eyed hero be...
6	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	Dramas	September 24, 2021	1993	TV-MA	125	On a photo shoot in Ghana, an American model s...
8	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy	Comedies	September 24, 2021	2021	PG-13	104	A woman adjusting to life after a loss contend...
10	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler	Dramas	September 23, 2021	2021	TV-MA	127	After most of her family is murdered in a terr...
...
1529022	s8404	Movie	The Look of Silence	Joshua Oppenheimer	El Negro Álvarez	International Movies	February 26, 2016	2014	PG-13	103	In a society terrorized into silence, the brot...
1688255	s6234	Movie	Barbecue	Matthew Salleh	El Negro Álvarez	Documentaries	August 15, 2017	2017	TV-MA	101	A blend of cultural nuance and mesmerizing tec...
1695579	s6234	Movie	Barbecue	Matthew Salleh	El Negro Álvarez	International Movies	August 15, 2017	2017	TV-MA	101	A blend of cultural nuance and mesmerizing tec...
1856466	s6234	Movie	Barbecue	Matthew Salleh	El Negro Álvarez	Documentaries	August 15, 2017	2017	TV-MA	101	A blend of cultural nuance and mesmerizing tec...
1863790	s6234	Movie	Barbecue	Matthew Salleh	El Negro Álvarez	International Movies	August 15, 2017	2017	TV-MA	101	A blend of cultural nuance and mesmerizing tec...

148286 rows × 12 columns

```
In [176]:  
  
# Extracting No. of Seasons  
  
shows_df['duration'] = shows_df['duration'].str.split().apply( lambda x: x[0] )  
shows_df['duration']
```

```
1      2  
2      1  
3      2  
4      1  
7      9  
..  
1237505  2  
1242798  2  
1246739  2  
1253068  2  
1258149  2  
Name: duration, Length: 61940, dtype: object
```


Concatinating Movies and TV Shows

```
In [177]:  
  
# Concatinating Movies and TV Shows  
  
df2 = pd.concat([movies_df, shows_df], axis=0)  
df2.reset_index(drop=True, inplace = True)  
df2['duration'] = df2['duration'].astype(int)  
df1 = df2  
df3 = df2  
df2
```

	show_id	type	title	director	cast	genre	date_added	release_year	rating	duration	description	cc
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Craig Sechler	Documentaries	September 25, 2021	2020	PG-13	90	As her father nears the end of his life, filmm...	Ur St
1	s7	Movie	My Little Pony: A New Generation	Robert Cullen	Vanessa Hudgens	Children & Family Movies	September 24, 2021	2021	PG	91	Equestria's divided. But a bright-eyed hero be...	Ur St
2	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba	Dramas	September 24, 2021	1993	TV-MA	125	On a photo shoot in Ghana, an American model s...	Ur St
3	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy	Comedies	September 24, 2021	2021	PG-13	104	A woman adjusting to life after a loss contend...	Ur St
4	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler	Dramas	September 23, 2021	2021	TV-MA	127	After most of her family is murdered in a terr...	G€
...
210221	s1507	TV Show	Shaun the Sheep	Jay Oliva	Richard Webber	Kids' TV	December 22, 2020	2016	TV-Y	2	Join freewheeling Shaun for barnyard misadvent...	Ire
210222	s1507	TV Show	Shaun the Sheep	Jay Oliva	Richard Webber	TV Comedies	December 22, 2020	2016	TV-Y	2	Join freewheeling Shaun for barnyard misadvent...	Ire
210223	s1507	TV Show	Shaun the Sheep	Jay Oliva	Jo Allen	British TV Shows	December 22, 2020	2016	TV-Y	2	Join freewheeling Shaun for barnyard misadvent...	Ire
210224	s1507	TV Show	Shaun the Sheep	Jay Oliva	Jo Allen	Kids' TV	December 22, 2020	2016	TV-Y	2	Join freewheeling Shaun for barnyard misadvent...	Ire
210225	s1507	TV Show	Shaun the Sheep	Jay Oliva	Jo Allen	TV Comedies	December 22, 2020	2016	TV-Y	2	Join freewheeling Shaun for barnyard misadvent...	Ire

210226 rows × 12 columns

In [178]:

converting Date_Added to date & time

```
df2['date_added'] = pd.to_datetime(df2['date_added'])
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 210226 entries, 0 to 210225
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   show_id         210226 non-null object
1   type            210226 non-null object
2   title           210226 non-null object
3   director        210226 non-null object
4   cast            210226 non-null object
5   genre           210226 non-null object
6   date_added      210226 non-null datetime64[ns]
7   release_year    210226 non-null int64
8   rating          210226 non-null object
9   duration        210226 non-null int64
10  description      210226 non-null object
11  country         210226 non-null object
dtypes: datetime64[ns](1), int64(2), object(9)
memory usage: 19.2+ MB
```

unnesting cast , director , country and gener(listed_in) is Done.

Also handeled duration column & converted to int64.

Also converted Date_Added to date & time

Hence data cleaning and preprocessing are done, now its ready to analyse. 🍏

In [179]:

df1.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 210226 entries, 0 to 210225
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   show_id         210226 non-null object
1   type            210226 non-null object
2   title           210226 non-null object
3   director        210226 non-null object
4   cast            210226 non-null object
5   genre           210226 non-null object
6   date_added      210226 non-null datetime64[ns]
7   release_year    210226 non-null int64
8   rating          210226 non-null object
9   duration        210226 non-null int64
10  description      210226 non-null object
11  country         210226 non-null object
dtypes: datetime64[ns](1), int64(2), object(9)
memory usage: 19.2+ MB
```

Analysis:-

business problem => which type of shows to produce and how to grow the business.

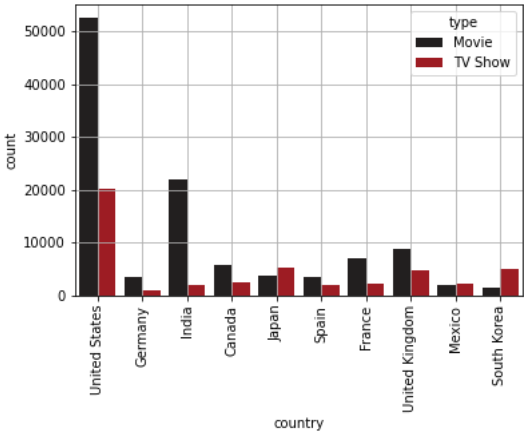
```
In [209]:
palette = ['#221F1F', '#B20710', '#E50914', '#F5F5F1']
sns.set_palette(palette)
sns.set_style(rc = {'axes.facecolor': "#FFFFFF"})
df1.describe(include='object')
```

	show_id	type	title	director	cast	genre	rating	description	country
count	210226	210226	210226	210226	210226	210226	210226	210226	210226
unique	8704	2	8704	4991	36428	42	14	8676	123
top	s7165	Movie	Kahlil Gibran's The Prophet	Jay Oliva	El Negro Álvarez	Dramas	TV-MA	A troubled young girl and her mother find sola...	United States
freq	700	148286	700	15904	448	29829	77779	700	72679

What type of content is available in different countries ?

```
In [214]:
con = df1['country'].value_counts().head(10)
con = con.reset_index()
data = df1[df1['country'].isin(con['index'])]

sns.countplot(data = data, x="country", hue="type") # Dodget Bar Chart
plt.xticks(rotation = 90)
plt.grid()
plt.show()
```



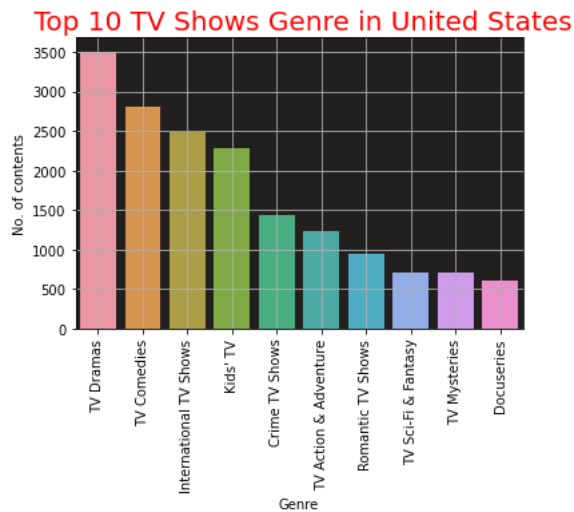
Comments: For top 10 countries majorly movies are available specially in United States and India.

There are some exceptions like south Korea and Japan.

which genre of movies are popular in usa and India ?

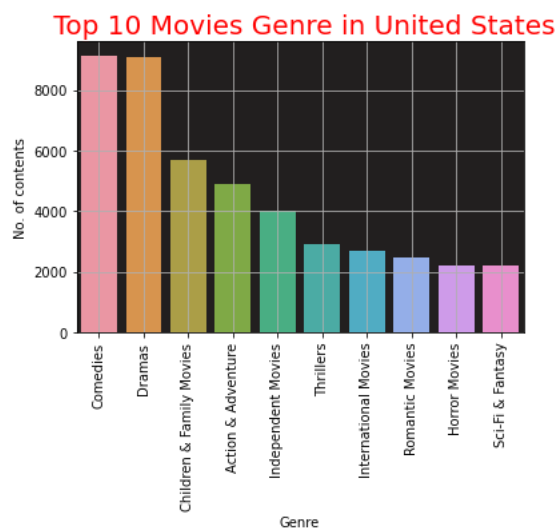
```
In [182]:
gen_us = df1[ (df1["country"]=="United States") & (df1["type"]=="TV Show") ]
gen = gen_us['genre'].value_counts().head(10)
gen = gen.reset_index()
gen.rename(columns={"index":"Genre" , "genre":"No. of contents"} , inplace = True)
gen

sns.barplot(data = gen, x="Genre", y = "No. of contents" ) # Boxplot
plt.xticks(rotation = 90)
plt.grid()
plt.title(" Top 10 TV Shows Genre in United States" , fontsize = 20 , color = "red" )
plt.show()
```



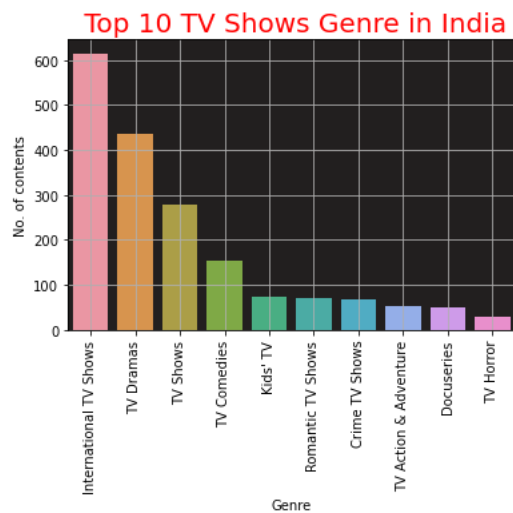
Comments: Mostly Comedie and Drama type shows are popular in USA.

```
In [183]:  
  
gen_us = df1[ (df1["country"]=="United States") & (df1["type"]=="Movie")]  
gen = gen_us['genre'].value_counts().head(10)  
gen = gen.reset_index()  
gen.rename(columns={"index":"Genre" , "genre":"No. of contents"} , inplace = True)  
gen_us  
  
sns.barplot(data = gen, x="Genre", y = "No. of contents" ) # Boxplot  
plt.xticks(rotation = 90)  
plt.grid()  
plt.title(" Top 10 Movies Genre in United States" , fontsize = 20 , color = "red" )  
plt.show()
```



Comments: Mostly comedy and Drama type movies are exceptionally good

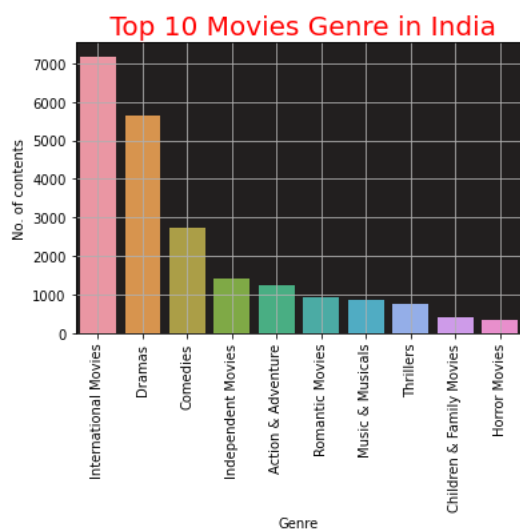
```
In [184]:  
  
ind_df = df1[ (df1["country"]=="India") & (df1["type"]=="TV Show")]  
ind = ind_df['genre'].value_counts().head(10)  
ind = ind.reset_index()  
ind.rename(columns={"index":"Genre" , "genre":"No. of contents"} , inplace = True)  
ind  
  
sns.barplot(data = ind, x="Genre", y = "No. of contents" ) # Boxplot  
plt.xticks(rotation = 90)  
plt.grid()  
plt.title(" Top 10 TV Shows Genre in India" , color = "red" , fontsize = 20)  
plt.show()
```



Comments: from 2008 - 2014 - Netflix has focused movies over TV shows But, In the 2008 the TV shows were added and after that directly 2013 was the next year where the TV shows were added so there is huge gap between these years but TV shows are tending in market right now so It is one of the factors to focus to grow the business

```
In [185]:
ind_df = df1[ (df1["country"]=="India") & (df1["type"]=="Movie")]
ind = ind_df['genre'].value_counts().head(10)
ind = ind.reset_index()
ind.rename(columns={"index":"Genre" , "genre":"No. of contents"} , inplace = True)
ind

sns.barplot(data = ind, x="Genre", y ="No. of contents" ) # Boxplot
plt.xticks(rotation = 90)
plt.grid()
plt.title(" Top 10 Movies Genre in India" , color = "red" , fontsize = 20)
plt.show()
```

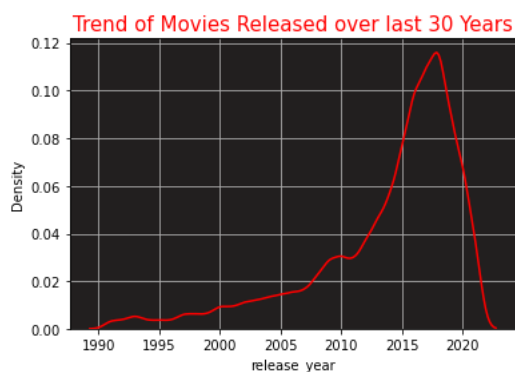


Comments: Top 2 genre among movies in India. - 'International Movies and Dramas'

How has the number of movies released per year changed over the last 20-30 years?

```
In [186]:
yyy = df2[ (df2['release_year'] > 1990) & (df2['type'] == "Movie") ]
yyy

sns.kdeplot(data=yyy, x="release_year", color = "red") # Kernel Density Estimation(KDE)
plt.grid()
plt.title('Trend of Movies Released over last 30 Years', color = "red", fontsize = 15)
plt.show()
```



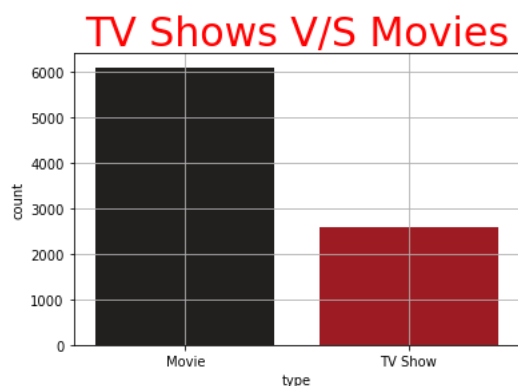
Comments: in the past 20 years number of movies per year was continuously increasing but the peak of higher released movies are in 2017-2018

but, due to covid it directly dropped to its min value in the year 2020-21

Comparison of tv shows vs. movies

```
In [213]:
data = df1
data = data.drop_duplicates("title")

sns.countplot(data = data, x="type") # Dodget Bar Chart
plt.title("TV Shows V/S Movies", fontsize = 30, color='red')
plt.grid()
plt.show()
```



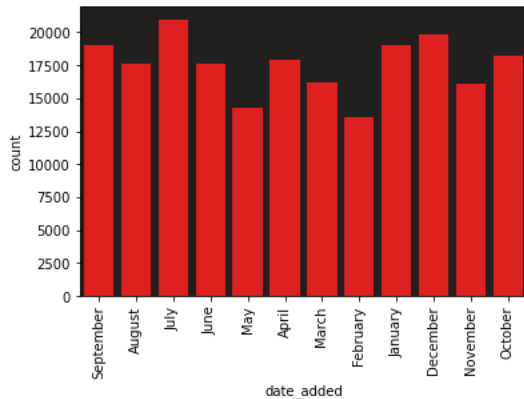
Comments: Movies count are much higher than the TV shows

What is the best time to launch a TV show?

```
In [188]:
data = df1["date_added"].dt.strftime('%B')
data = data.reset_index()

sns.countplot(data = data, x="date_added", color='red') # Dodget Bar Chart
plt.title("Months in TV Shows or Movies were Added", fontsize = 20, color='red')
plt.xticks(rotation = 90)
plt.show()
```

Months in TV Shows or Movies were Added



Comments: **Top 2 months for adding higher number of movies or TV shows: 'July and December'**

Analysis of actors/directors of different types of shows/movies.

Top 5 Directors in who worked in maximum number of movies & TV Shows

```
In [189]:
movies = df1[df1['type'] == "Movie"]
tv = df1[df1['type'] == "TV Show"]

movies = movies.drop_duplicates(['title', 'director'])
tv = tv.drop_duplicates(['title', 'director'])
```

```
In [190]:
movies['director'].value_counts().head(5)
```

```
Jan Suter      31
Raúl Campos    29
Marcus Raboy    22
Rajiv Chilaka  22
Richard Bailey  18
Name: director, dtype: int64
```

Comments: top 5 movie directors having the range from 18-31 for maximum number of movies where they worked

```
In [191]:
tv['director'].value_counts().head(5)
```

```
Jay Oliva          659
Tony Collingwood   198
Mat King           101
Ehtesham Uddin     89
Michael Simon      87
Name: director, dtype: int64
```

Comments: top 5 movie directors having the range from 87-659 for maximum number of tv shows where they worked

```
In [192]:
### Top 5 Actors in wh worked in mximum number of movies & TV Show.
```

```
In [193]:
movies = df1[df1['type'] == "Movie"]
tv = df1[df1['type'] == "TV Show"]

movies = movies.drop_duplicates(['title', "cast"])
tv = tv.drop_duplicates(['title', "cast"])
```

```
In [194]:
movies['cast'].value_counts().head(5)
```

```
Craig Sechler      136
El Negro Álvarez   70
Anupam Kher        42
Shah Rukh Khan     35
Ben Ferencz        33
Name: cast, dtype: int64
```

Comments: top 5 Movie Cast having the range from 33-136 for maximum number of movies where they worked

```
In [195]:
tv['cast'].value_counts().head(5)
```

```
Morgan Freeman     36
R. Kelly           32
David Attenborough  29
Takahiro Sakurai    25
Marie Kondo         21
Name: cast, dtype: int64
```

Comments: top 5 TV Cast having the range from 21-36 for maximum number of TV shows where they worked

Rating of Movies and TV shows in netflix in United States.

```
In [196]:
movies_us = df1[(df1['type'] == "Movie") & (df1['country'] == "United States")]
tv_us = df1[(df1['type'] == "TV Show") & (df1['country'] == "United States")]

movies_us = movies_us.drop_duplicates(['title'])
tv_us = tv_us.drop_duplicates(['title'])
```

```
In [197]:
movies_us['rating'].value_counts().head(5)
```

```
TV-MA    838
R         670
PG-13     440
TV-14     376
PG        248
Name: rating, dtype: int64
```

Comments:top 5 ratings for movies in United States :

TV-MA :(2058) TV-14 :(1423) R :(797) TV-PG :(539) PG-13 :(490)

```
In [198]:
tv_us['rating'].value_counts().head(5)
```

```
TV-MA    486
TV-14     311
TV-PG     161
TV-Y7     128
TV-Y       97
Name: rating, dtype: int64
```

Comments: **ANALYSIS : top 5 ratings for TV shows in United States :

TV-MA :(1116) TV-14 :(716) TV-PG :(306) TV-Y7 :(189) TV-Y :(168) *

Rating of Movies and TV shows in netflix in India.

```
In [199]:
movies_ind = df1[(df1['type'] == "Movie") & (df1['country'] == "India")]
tv_ind = df1[(df1['type'] == "TV Show") & (df1['country'] == "India")]

movies_ind = movies_ind.drop_duplicates(['title'])
tv_ind = tv_ind.drop_duplicates(['title'])
```

```
In [200]:
movies_ind['rating'].value_counts().head(5)
```

```
TV-14    559
TV-MA    235
TV-PG    133
TV-Y7     26
PG-13     11
Name: rating, dtype: int64
```

Comments: **ANALYSIS : top 5 ratings for TV Movies in India :

TV-MA :(2058) TV-14 :(1423) R :(797) TV-PG :(539) PG-13 :(490) **

```
In [201]:
tv_ind['rating'].value_counts().head(5)
```

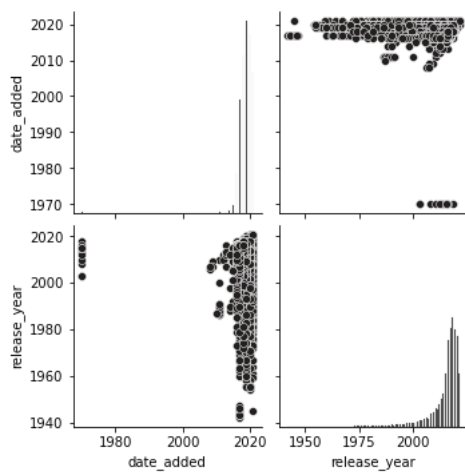
```
TV-MA    34
TV-14     25
TV-PG     11
TV-Y       6
TV-Y7      5
Name: rating, dtype: int64
```

Comments: **ANALYSIS : top 5 ratings for TV shows in India :

TV-MA :(1116) TV-14 :(716) TV-PG :(306) TV-Y7 :(189) TV-Y :(168) **

Relation ship between date added and release year

```
In [212]:  
data = df1[["date_added", "release_year", 'title']]  
data = data.drop_duplicates('title')  
data = data[["date_added", "release_year"]]  
data["date_added"] = data["date_added"].dt.year  
data["date_added"] = data["date_added"].astype(int)  
sns.pairplot(data)  
plt.show()
```

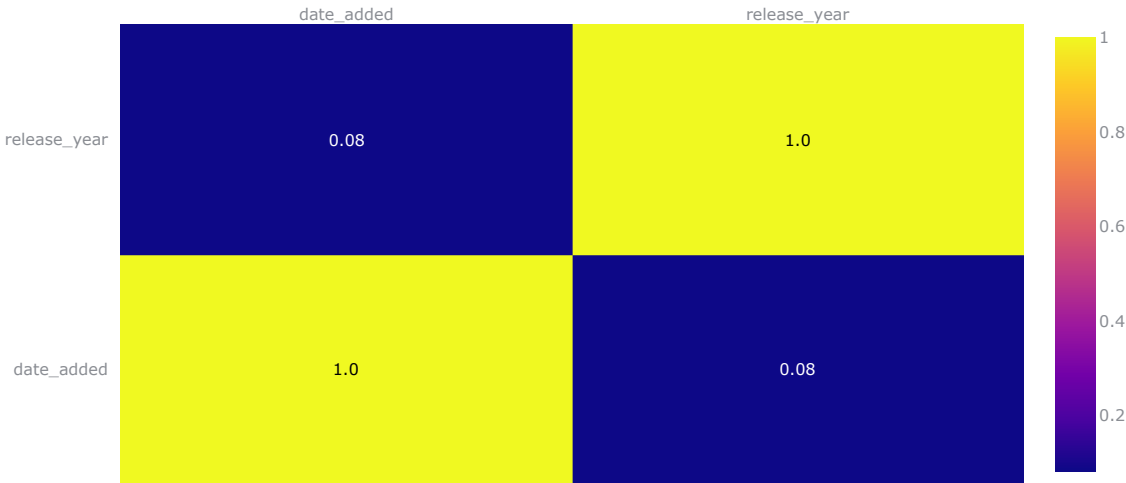


```
In [203]: data.info()
import plotly.figure_factory as ff

corrs = data.corr()
fig_heatmap = ff.create_annotated_heatmap(
    z=corrs.values,
    x=list(corrs.columns),
    y=list(corrs.index),
    annotation_text=corrs.round(2).values,
    showscale=True)
fig_heatmap.update_layout(title= 'Correlation of whole Data',
                           plot_bgcolor='#2d3035', paper_bgcolor='#2d3035',
                           title_font=dict(size=25, color='#a5a7ab', family="Mul", sans-serif"),
                           font=dict(color='#8a8d93'))
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8704 entries, 0 to 150880
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   date_added   8704 non-null   int64
1   release_year 8704 non-null   int64
dtypes: int64(2)
memory usage: 462.0 KB
```

Correlation of whole Data



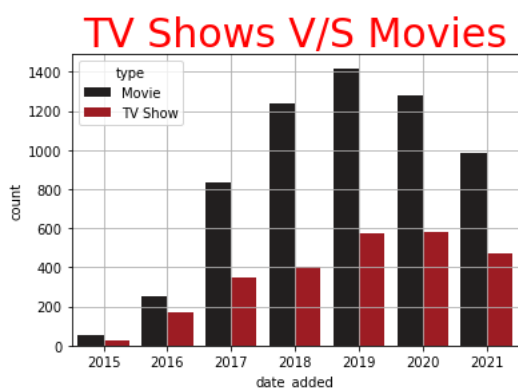
Comments: The above analysis between date added and release year shows that netflix releases

TV shows and movies as soon as they are released.

Does Netflix has more focus on TV Shows than movies in recent years

```
In [211]:
data = df1[ (df1['date_added'].dt.year > 2014) ]
data['date_added'] = data['date_added'].dt.year
data = data.drop_duplicates("title")
data

sns.countplot(data = data, x='date_added' , hue="type") # Dodget Bar Chart
plt.title("TV Shows V/S Movies" , fontsize = 30 , color='red')
plt.grid()
plt.show()
```



Comments: from 2015 - 2021 - Netflix has focused movies over TV shows where as TV shows seems in trending

top 10 popular TV shows

```
In [205]:
tv = df1[df1['type'] == "TV Show"]
tv = tv.drop_duplicates(['title'])
# tv = tv[['title', 'duration']]
tv[['title', 'duration']].sort_values(by = 'duration' , ascending = False).head(10).reset_index(drop = True)
```

	title	duration
0	Grey's Anatomy	17
1	Supernatural	15
2	NCIS	15
3	Heartland	13
4	COMEDIANS of the world	13
5	Red vs. Blue	13
6	Criminal Minds	12
7	Trailer Park Boys	12
8	Cheers	11
9	Frasier	11

Comments: top 10 TV shows has range of total seasons are (11 to 17) where as top3 are having total of 17 ,15,15 seasons respectively

In [206]:

```
top1 = df[df['title'] == "Grey's Anatomy"]
top1
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
548	s549	TV Show	Grey's Anatomy	NaN	Ellen Pompeo, Sandra Oh, Katherine Heigl, Just...	United States	July 3, 2021	2020	TV-14	17 Seasons	Romantic TV Shows, TV Dramas	Intern (and eventual resident) Meredith Grey f...

Top 1 Show has the maximum number of 17 seasons which indicates that its a very popular show from United States and has Rating:TV-14. This TV show falls in Romantic & Dramas Genre

In [207]:

```
top2 = df[df['title'] == "Supernatural"]
top2
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descriptio
2423	s2424	TV Show	Supernatural	Phil Sgriccia	Jared Padalecki, Jensen Ackles, Mark Sheppard,...	United States, Canada	June 5, 2020	2019	TV-14	15 Seasons	Classic & Cult TV, TV Action & Adventure, TV H...	Siblings Dean and Sam crisscross the country, ...

Top 2nd Show has the maximum number of 15 seasons which indicates that its very popular show from United States & Cananda and has Rating:TV-14. This TV show falls in Classic & Cult TV, TV Action & Adventure and this show is directed by Phil Sgriccia

In [208]:

```
top3 = df[df['title'] == "NCIS"]
top3
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
4798	s4799	TV Show	NCIS	NaN	Mark Harmon, Michael Weatherly, Pauley Perrett...	United States	July 1, 2018	2017	TV-14	15 Seasons	Crime TV Shows, TV Dramas, TV Mysteries	Follow the quirky agents of the NCIS – the Nav...

Top 3rd Show has the maximum number of 15 seasons which indicates that its very popular show from United States and has Rating:TV-14. This TV show falls in Crime TV Shows, TV Dramas, TV Mysteries

Business Insights

- 1) In the List of Top 10 countries like United States and India having majority of movies, compared to South Asian countries like south Koria and Japan
- 2) Top 2 genre among movies in USA - 'Comedies and Dramas' are top 2 popular in USA
- 3) Top 2 genre among TV Shows in USA : 'TV Dramas and TV Comedies'
- 4) Top 2 genre among movies in India. - 'International Movies and Dramas'
- 5) Top 2 genre among TV shows in India. - 'International TV Shows and TV Dramas'
- 6) in the past 20 years number of movies per year was continuously incresing but the peak of higher released movies are in 2017-2018 but, due to covid it directly dropped to its min value in the year 2020-21
- 7) Top 2 months for adding higher number of movies or TV shows: 'July and December'
- 8) Top 5 movie directors having the range from 18-31 for maximum number of movies where they worked
- 9) Top 5 movie directors having the range from 87-659 for maximum number of tv shows where they worked

- 10) Top 5 Movie Cast having the range from 33-136 for maximum number of movies where they worked¶
- 11) Top 5 TV Cast having the range from 21-36 for maximum number of TV shows where they worked¶
- 12) Top 5 ratings for movies in United States : TV-MA : (2058) TV-14 : (1423) R : (797) TV-PG : (539) PG-13 : (490)
- 13) Top 5 ratings for TV shows in United States : TV-MA : (1116) TV-14 : (716) TV-PG : (306) TV-Y7 : (189) TV-Y : (168)
- 14) Top 5 ratings for TV Movies in India : TV-MA : (2058) TV-14 : (1423) R : (797) TV-PG : (539) PG-13 : (490)
- 15) Top 5 ratings for TV shows in India : TV-MA : (1116) TV-14 : (716) TV-PG : (306) TV-Y7 : (189) TV-Y : (168)
- 16) The above analysis between date added and release year shows that netflix releases TV shows and movies as soon as they are released.
- 17) From 2015 - 2021 - Netflix has focused movies over TV shows where as TV shows seems in trending
- 18) From 2008 - 2014 - Netflix has focused movies over TV shows But, In the 2008 the TV shows were added and after that directly 2013 was the next year where the TV shows were added so there is huge gap between these years but TV shows are trending in market right now so It is one of the factors to focus to grow the business
- 19) Top 10 TV shows has range of total seasons are (11 to 17) where as top3 are having total of 17 ,15,15 seasons respectively
- 20) Top 1 Show has the maximum number of 17 seasons which indicates that its a very popular show from United States and has Rating:TV-14. This TV show falls in Romantic & Dramas Genre
- 21) Top 2nd Show has the maximum number of 15 seasons which indicates that its very popular show from United States & Canada and has Rating:TV-14. This TV show falls in Classic & Cult TV, TV Action & Adventure and this show is directed by Phil Sgriccia
- 22) Top 3rd Show has the maximum number of 15 seasons which indicates that its very popular show from United States and has Rating:TV-14. This TV show falls in Crime TV Shows, TV Dramas, TV Mysteries

Recommendations

- Netflix needs to have different strategy in terms of producing content for western countries and south asian countries.
- Drama is the most popular genre worldwide.So, Netflix can always rely on Drama type content for increasing market share and its growth in different countries.
- There was slight drop in content recently due to covid but company was performing well.
- Always try to release content during festive seasons or holidays like in the month of july and december with respect to location.
- Following are the top 5 experienced directors in the field of movies: Jan Suter Raúl Campos Marcus Raboy Rajiv Chilaka Richard Bailey
- So, while creating a new content these are the best ones to select in terms of experience.
- If nextflix wants low budget and young talent then low experienced cast are the most suitable.
- TV-MA is highly rated content.
- Netflix made a great market in movies but there is a good market scope for TV shows as they are trending. The TV shows genre in demand are :Drama and comedies.So netflix can produce more
- content in these 2 genre with rating TV-MA. for reference there are few TV shows who has more than 10 seasons.

Type *Markdown* and LaTeX: α^2