

In [1]:

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

In [2]:

```
df.shape
```

Out[2]:

```
(8807, 12)
```

In [3]:

```
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 [ ]:

In [4]:

```
df.describe()
```

Out[4]:

```
release_year
```

---

```
count    8807.000000
mean     2014.180198
std      8.819312
min     1925.000000
25%    2013.000000
50%    2017.000000
75%    2019.000000
max     2021.000000
```

In [5]:

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

Out[5]:

	show_id	type	title	director	cast	country	date_added	rating	duration
count	8807	8807	8807	6173	7982	7976	8797	8803	8804
unique	8807	2	8807	4528	7692	748	1767	17	220
top	s1	Movie	Dick Johnson Is Dead	Rajiv Chilaka	David Attenborough	United States	January 1, 2020	TV-MA	1 Season
freq	1	6131	1	19	19	2818	109	3207	1793

In [6]:

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

Out[6]:

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

In [7]:

```
df.loc[df['rating'].isin(['74 min','84 min','66 min']), 'duration']=df['rating']
```

In [8]:

```
df.loc[df['rating'].isin(['74 min','84 min','66 min']), 'rating']=np.nan
```

In [9]:

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

Out[9]:

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

In [10]:

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

Out[10]:

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

In [11]:

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

Out[11]:

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	1
Name: director, Length: 4528, dtype: int64	

In [12]:

df

Out[12]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009

show_id	type	title	director	cast	country	date_added	release_year	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	March 2, 2019	2015

8807 rows × 12 columns

In [13]:

```
df['date_added']=pd.to_datetime(df['date_added'])
```

In [14]:

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

```
'United Kingdom, Canada, United States', ', South Korea',
'Ireland', 'India, Nepal',
'New Zealand, Australia, France, United States', 'Italy',
'Italy, Brazil, Greece', 'Argentina', 'Jordan', 'Colombia',
'United States, Japan', 'Belgium, United Kingdom',
'Switzerland, United Kingdom, Australia', 'Israel, United States',
'Canada, United States', 'Brazil', 'Argentina, Spain', 'Taiwan',
'United States, Nigeria', 'Bulgaria, United States',
'Spain, United Kingdom, United States', 'United States, China',
'United States, France',
'Spain, France, United Kingdom, United States',
', France, Algeria', 'Poland', 'Germany',
'France, Israel, Germany, United States, United Kingdom',

'New Zealand', 'Saudi Arabia', 'Thailand', 'Indonesia',
'Egypt, Denmark, Germany', 'United States, Switzerland',
'Hong Kong, Canada, United States', 'Kuwait, United States',
'France, Canada, United States, Spain',
'France, Netherlands, Singapore', 'France, Belgium',
'Ireland, United States, United Kingdom', 'Egypt', 'Malaysia',
```

In [15]:

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

Out[15]:

```
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 [16]:

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

Out[16]:

```
array(['Movie', 'TV Show'], dtype=object)
```

In [ ]:

In [17]:

df

Out[17]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	2021-09-24	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	2021-09-24	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021-09-24	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	2019-11-20	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	2019-11-01	2009

show_id	type	title	director	cast	country	date_added	release_year	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	2020-01-11	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	2019-03-02	2015

8807 rows × 12 columns

In [18]:

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   datetime64[ns]
 7   release_year 8807 non-null   int64  
 8   rating      8800 non-null   object 
 9   duration    8807 non-null   object 
 10  listed_in   8807 non-null   object 
 11  description 8807 non-null   object 
dtypes: datetime64[ns](1), int64(1), object(10)
memory usage: 825.8+ KB

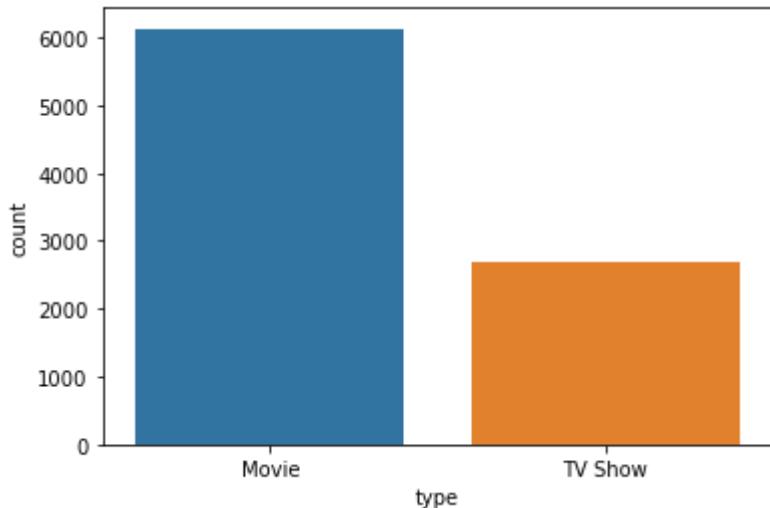
```

In [19]:

```
sns.countplot(x='type',data=df)
# Bar plot for the count for number of movies and TV Show
# Insight- Movies are 100% more than TV shows on Netflix
```

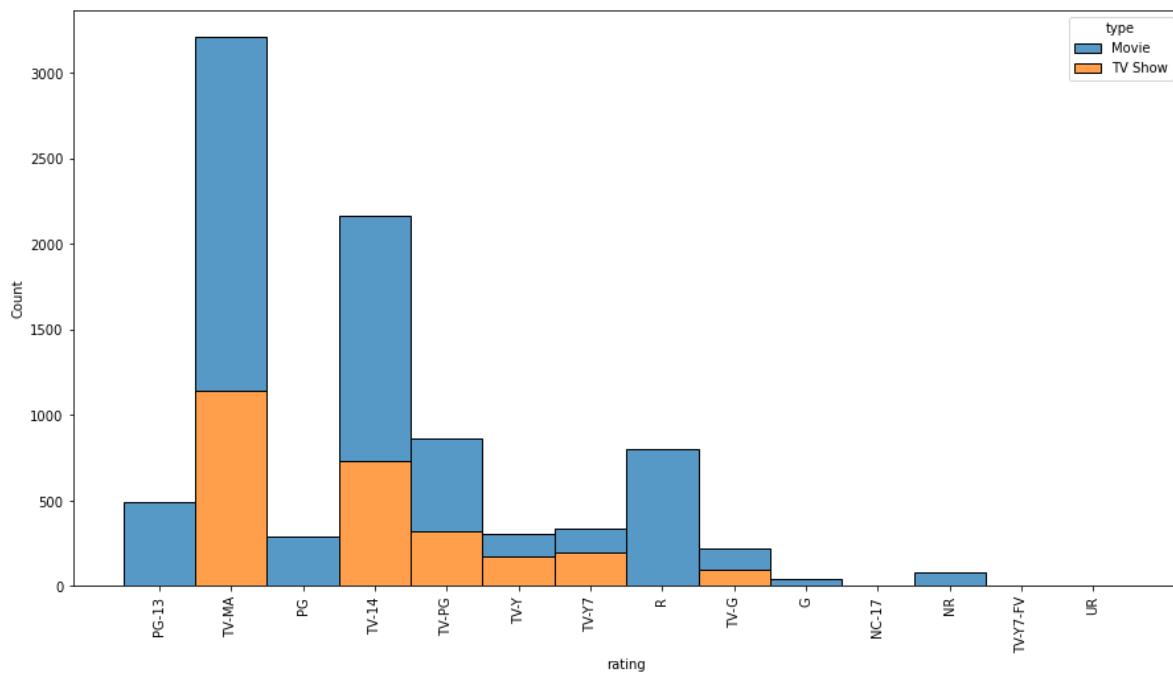
Out[19]:

&lt;AxesSubplot:xlabel='type', ylabel='count'&gt;



In [20]:

```
sns.histplot(x='rating',data=df,hue='type',multiple='stack')
plt.xticks(rotation=90)
plt.gcf().set_size_inches(15, 8)
plt.show()
# Hist plot for ratings vs count and type
# Insight - People like rating 'TV-MA', 'TV-14', 'TV-PG' and 'R'.
# There are no TV shows in 'PG-13' and 'R'.
```

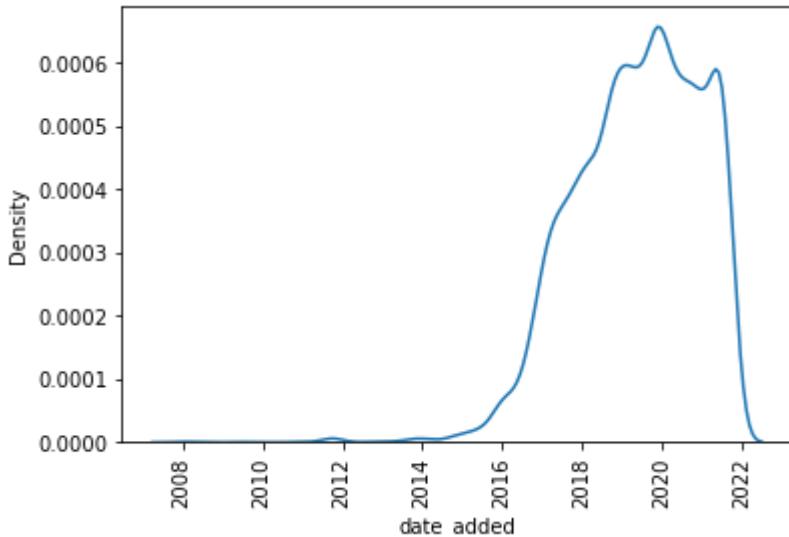


In [21]:

```

sns.kdeplot(x='date_added', data=df)
plt.xticks(rotation=90)
plt.show()
# Insight - There is a sudden increase in number of movies /TV shows that are released on N
# Slight dip in movies/TV shows released in 2020 year
# The Trend Looks increasing in the forth comming years

```

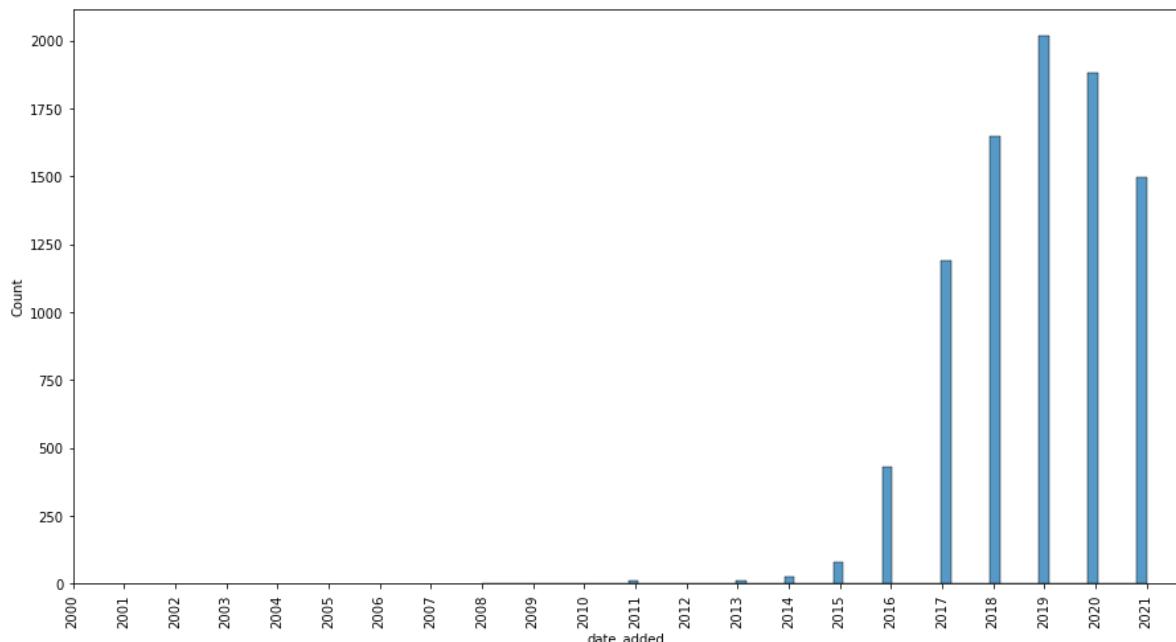


In [22]:

```

sns.histplot(x=df['date_added'].dt.year)
plt.xticks(np.arange(2000,2022),rotation=90)
plt.gcf().set_size_inches(15, 8)
plt.show()
# Insight - There is an increase in the number of movies exponentially from 2013

```

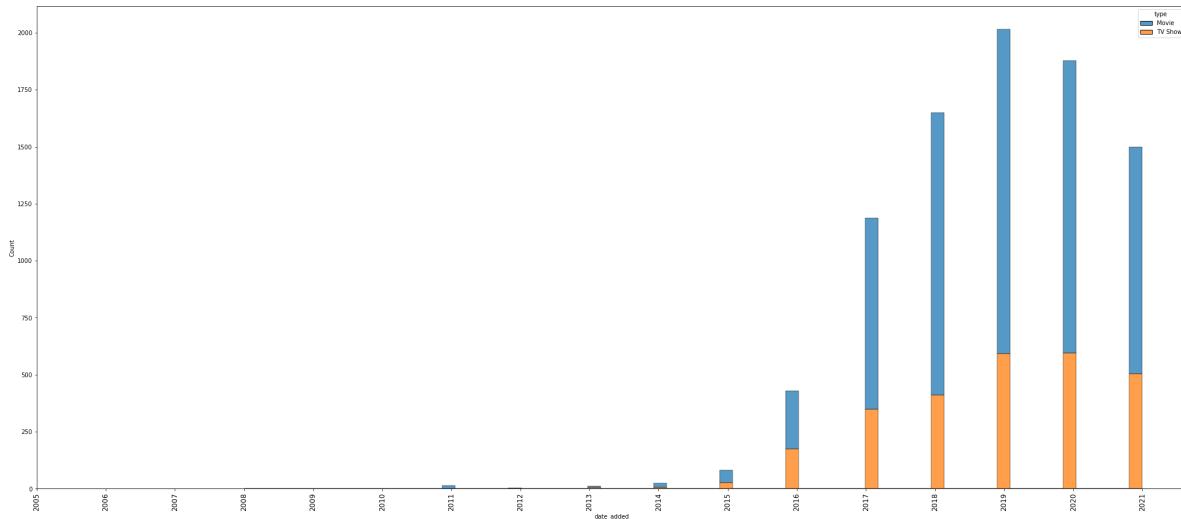


In [23]:

```

sns.histplot(x=df['date_added'].dt.year,hue=df['type'],multiple="stack")
plt.xticks(np.arange(2005,2022),rotation=90,fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()
# Insight - There is an increase in the number of movies and TV Show exponentially from 2011

```

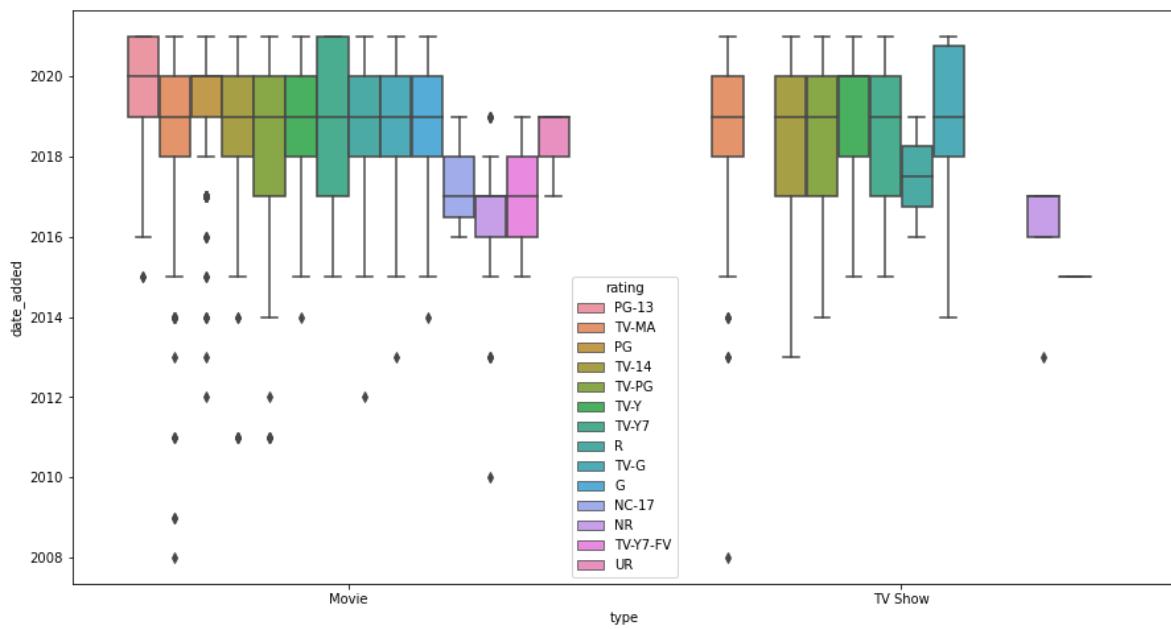


In [24]:

```

sns.boxplot(x=df['type'],y=df['date_added'].dt.year,hue=df['rating'])
plt.gcf().set_size_inches(15, 8)
plt.show()
# Insight- Outliner points are shown in boxplot which needs to remove from our analysis
# Movie released range for each 'rating' are shown in boxplot for movie and TV show

```



In [25]:

```
countrysplit=df['country'].apply(lambda x: str(x).split(', ')).tolist()
countrysplits=pd.DataFrame(countrysplit,index=df['title'])
countrysplits=countrysplits.stack().reset_index()
countrysplits.drop('level_1',axis=1,inplace=True)
countrysplits.rename({0:'listed_coun'},axis=1,inplace=True)
new=df.merge(countrysplits, left_on='title', right_on='title', how='left')
# Splitting country column
```

In [26]:

```
new['listed_coun'].value_counts()
# High number of movies are released in US, India, UK and Canada.
# Below table shows the number of movies released in each country
```

Out[26]:

United States	3689
India	1046
nan	831
United Kingdom	804
Canada	445
...	
Bermuda	1
Ecuador	1
Armenia	1
Mongolia	1
Montenegro	1

Name: listed\_coun, Length: 128, dtype: int64

In [27]:

```
new.drop(['country','description'],axis=1,inplace=True)
```

In [28]:

new

Out[28]:

	show_id	type	title	director	cast	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	2021-09-25	2020	PG-13	90
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021	TV-MA	Sea
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	2021-09-24	2021	TV-MA	Se
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	2021-09-24	2021	TV-MA	Se
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	2021-09-24	2021	TV-MA	Sea
...	...	...	...	...	...	...	...	...	...
10840	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	2019-11-20	2007	R	158
10841	s8804	TV Show	Zombie Dumb	NaN	NaN	2019-07-01	2018	TV-Y7	Sea
10842	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	2019-11-01	2009	R	88
10843	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	2020-01-11	2006	PG	88

show_id	type	title	director	cast	date_added	release_year	rating	duration
10844	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	2019-03-02	2015	TV-14

10845 rows × 11 columns



In [29]:

```
ind=new.loc[new['listed_coun']==nan].index
new_1=new.drop(ind,axis=0)
```

In [ ]:



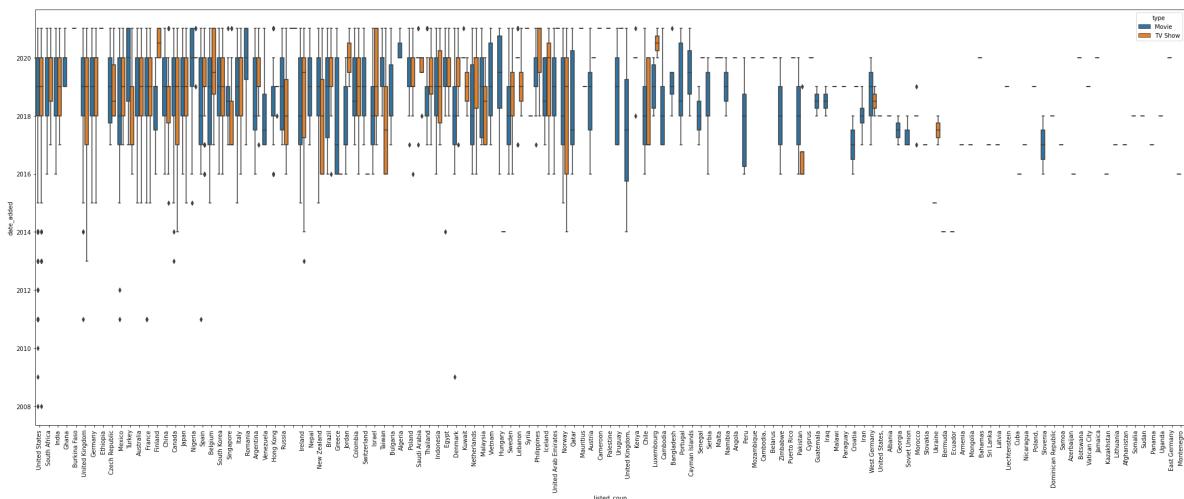
In [ ]:



In [30]:

```
sns.boxplot(x=new_1['listed_coun'],y=new_1['date_added'].dt.year,hue=new_1['type'])
plt.xticks(rotation=90)
plt.gcf().set_size_inches(35, 13)
plt.show()
```

# Barplot shows all the country wise movie released range along with type



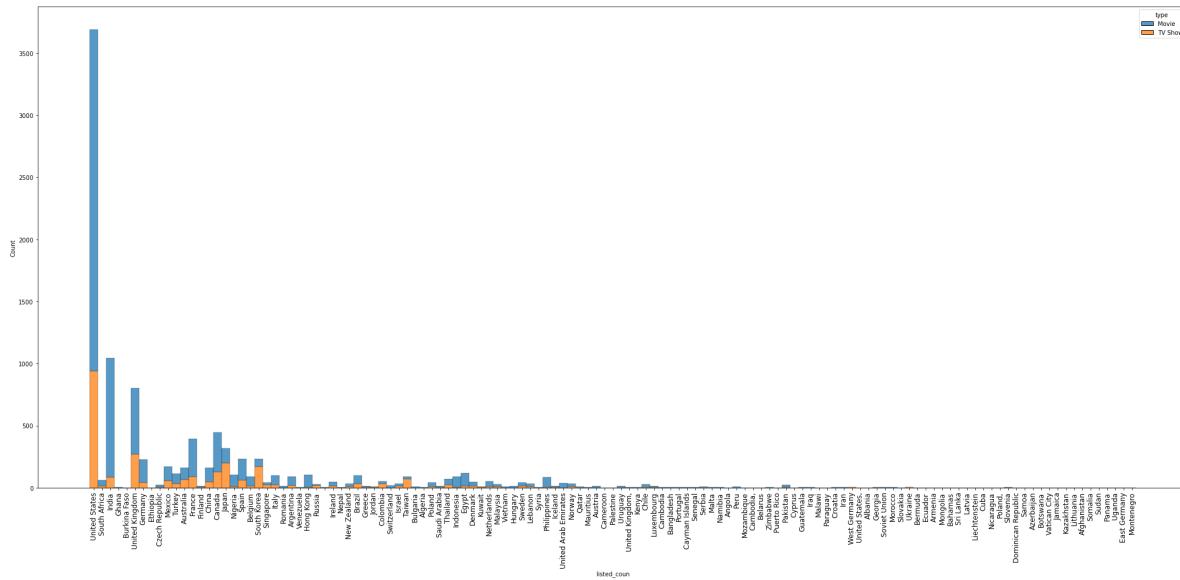
In [31]:

```

sns.histplot(x=new_1['listed_coun'], hue=new_1['type'], multiple="stack")
plt.xticks(rotation=90, fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()

# In US people like more movies than TV shows
# Japan , South Korea and Taiwan Like more TV shows than movies

```



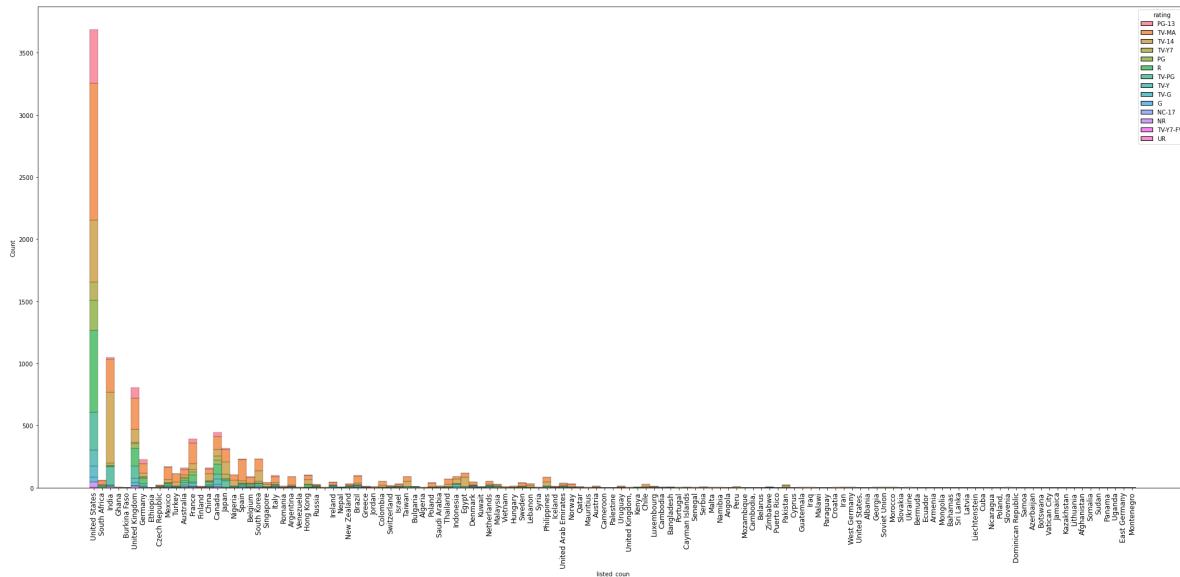
In [32]:

```

sns.histplot(x=new_1['listed_coun'], hue=new_1['rating'], multiple="stack")
plt.xticks(rotation=90, fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()

# Plot shows country wise frequency of each unique rating
# From seeing th plot we can conclude which type of rating each country like or dislike more

```



In [ ]:

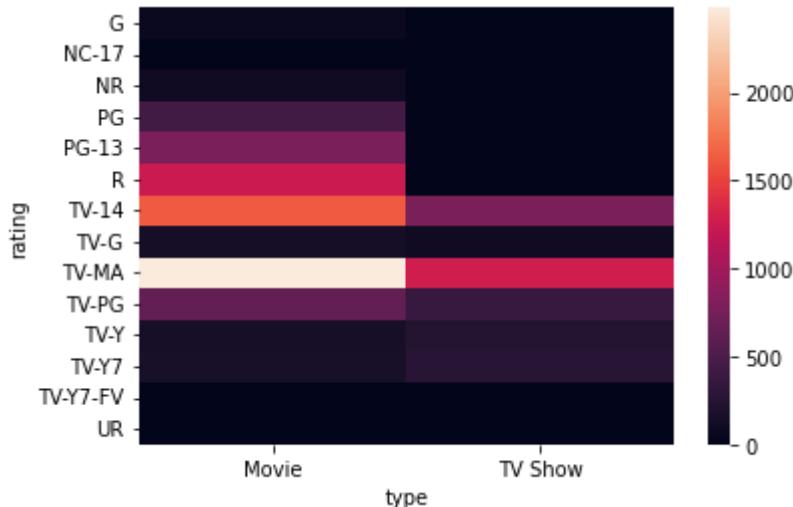
In [ ]:

In [33]:

```
sns.heatmap(pd.crosstab(new['rating'],new['type']))
# Heatmap of Movie and TV show along with rating
# In movies people like TV-14,TV-MA,TV-PG, R, PG-13, PG the most
# In TV shows people like TV-14,TV-MA,TV-Y the most
```

Out[33]:

```
<AxesSubplot:xlabel='type', ylabel='rating'>
```



In [34]:

```
new1=pd.crosstab(new['listed_coun'],new['rating']).reset_index()
```

In [35]:

```
new1=new1.reset_index()
```

In [36]:

```
new1=new1.set_index('index').reset_index()
```

In [37]:

new1

Out[37]:

rating	index	listed_coun	G	NC-17	NR	PG	PG-13	R	TV-14	TV-G	TV-MA	TV-PG	TV-Y	TV-Y7	TV-Y7-FV	UR
0	0		0	0	0	0	0	0	1	0	1	0	0	0	0	0
1	1	Afghanistan	0	0	0	0	0	0	0	0	1	0	0	0	0	0
2	2	Albania	0	0	0	0	0	0	0	0	1	0	0	0	0	0
3	3	Algeria	0	0	0	0	0	0	1	0	2	0	0	0	0	0
4	4	Angola	0	0	0	0	0	0	0	0	1	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
123	123	Venezuela	0	0	1	0	0	0	2	0	1	0	0	0	0	0
124	124	Vietnam	0	0	0	0	0	0	3	1	3	0	0	0	0	0
125	125	West Germany	1	0	0	0	0	0	1	0	3	0	0	0	0	0
126	126	Zimbabwe	0	0	1	0	0	0	0	1	1	0	0	0	0	0
127	127	nan	0	0	0	6	8	11	230	30	276	90	80	98	1	0

128 rows × 16 columns

In [38]:

new1=new1.set\_index('index')

In [39]:

new1

Out[39]:

rating	listed_coun	G	NC-17	NR	PG	PG-13	R	TV-14	TV-G	TV-MA	TV-PG	TV-Y	TV-Y7	TV-Y7-FV	UR
index															
0		0	0	0	0	0	0	1	0	1	0	0	0	0	0
1	Afghanistan	0	0	0	0	0	0	0	0	1	0	0	0	0	0
2	Albania	0	0	0	0	0	0	0	0	1	0	0	0	0	0
3	Algeria	0	0	0	0	0	0	1	0	2	0	0	0	0	0
4	Angola	0	0	0	0	0	0	0	0	1	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
123	Venezuela	0	0	1	0	0	0	2	0	1	0	0	0	0	0
124	Vietnam	0	0	0	0	0	0	3	1	3	0	0	0	0	0
125	West Germany	1	0	0	0	0	0	1	0	3	0	0	0	0	0
126	Zimbabwe	0	0	1	0	0	0	0	1	1	0	0	0	0	0
127	nan	0	0	0	6	8	11	230	30	276	90	80	98	1	0

128 rows × 15 columns

In [40]:

new1.columns

Out[40]:

```
Index(['listed_coun', 'G', 'NC-17', 'NR', 'PG', 'PG-13', 'R', 'TV-14', 'TV-G',
       'TV-MA', 'TV-PG', 'TV-Y', 'TV-Y7', 'TV-Y7-FV', 'UR'],
      dtype='object', name='rating')
```

In [41]:

new1.drop([0,127],inplace=True)

In [42]:

new1

Out[42]:

rating	listed_coun	G	NC-17	NR	PG	PG-13	R	TV-14	TV-G	TV-MA	TV-PG	TV-Y	TV-Y7	TV-Y7-FV	UR
index															
1	Afghanistan	0	0	0	0	0	0	0	0	1	0	0	0	0	0
2	Albania	0	0	0	0	0	0	0	0	1	0	0	0	0	0
3	Algeria	0	0	0	0	0	0	1	0	2	0	0	0	0	0
4	Angola	0	0	0	0	0	0	0	0	1	0	0	0	0	0
5	Argentina	1	0	2	1	0	2	10	2	67	4	2	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
122	Vatican City	0	0	0	1	0	0	0	0	0	0	0	0	0	0
123	Venezuela	0	0	1	0	0	0	2	0	1	0	0	0	0	0
124	Vietnam	0	0	0	0	0	0	3	1	3	0	0	0	0	0
125	West Germany	1	0	0	0	0	0	1	0	3	0	0	0	0	0
126	Zimbabwe	0	0	1	0	0	0	0	1	1	0	0	0	0	0

126 rows × 15 columns

In [43]:

new1=new1.set\_index('listed\_coun')

In [44]:

new1

Out[44]:

	rating	G	NC-17	NR	PG	PG-13	R	TV-14	TV-G	TV-MA	TV-PG	TV-Y	TV-Y7	TV-Y7-FV	UR
listed_coun															
Afghanistan	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Albania	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Algeria	0	0	0	0	0	0	1	0	2	0	0	0	0	0	0
Angola	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Argentina	1	0	2	1	0	2	10	2	67	4	2	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Vatican City	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Venezuela	0	0	1	0	0	0	2	0	1	0	0	0	0	0	0
Vietnam	0	0	0	0	0	0	3	1	3	0	0	0	0	0	0
West Germany	1	0	0	0	0	0	1	0	3	0	0	0	0	0	0
Zimbabwe	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0

126 rows × 14 columns

In [45]:

new2=new1.idxmax(axis=1).reset\_index()

In [ ]:

In [46]:

new2.rename({0:'type'},axis=1,inplace=True)

In [47]:

```
new2  
# Shows the most Liked rating in each country
```

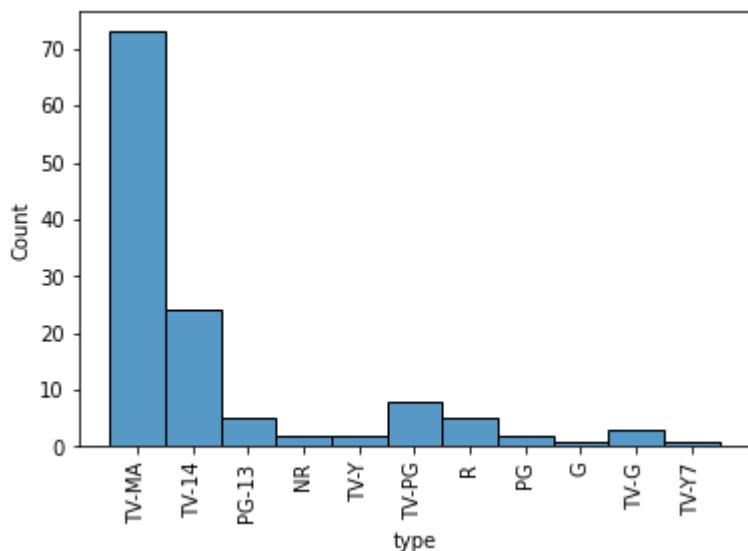
Out[47]:

	listed_coun	type
0	Afghanistan	TV-MA
1	Albania	TV-MA
2	Algeria	TV-MA
3	Angola	TV-MA
4	Argentina	TV-MA
...	...	...
121	Vatican City	PG
122	Venezuela	TV-14
123	Vietnam	TV-14
124	West Germany	TV-MA
125	Zimbabwe	NR

126 rows × 2 columns

In [48]:

```
sns.histplot(x=new2['type'])  
plt.xticks(rotation=90)  
plt.show()  
# More than 70 countries like 'TV-MA' the most, more than 24 countries like 'TV-14' the mos
```

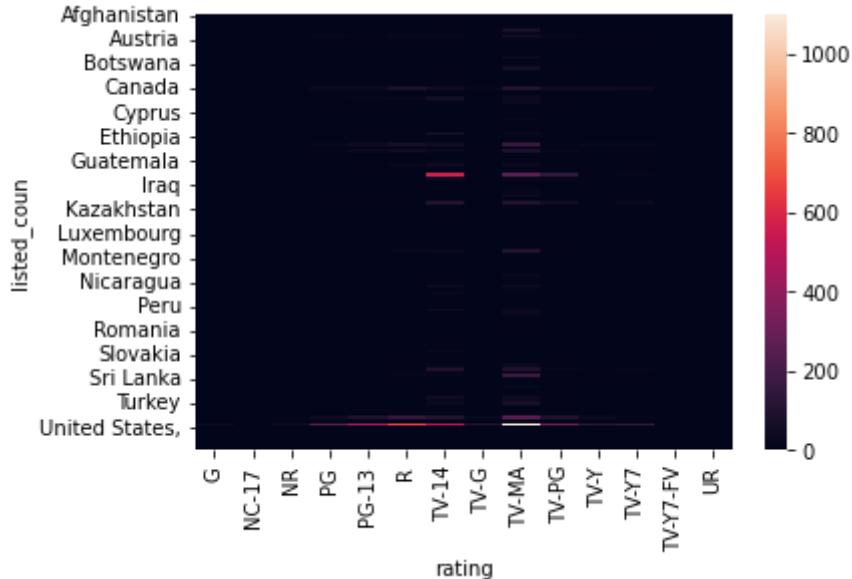


In [49]:

```
sns.heatmap(new1)
```

Out[49]:

```
<AxesSubplot:xlabel='rating', ylabel='listed_coun'>
```



In [50]:

new

Out[50]:

	show_id	type	title	director	cast	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	2021-09-25	2020	PG-13
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021	TV-MA
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	2021-09-24	2021	TV-MA
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	2021-09-24	2021	TV-MA
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	2021-09-24	2021	TV-MA
...	...	...	...	...	...	...	...	...
10840	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	2019-11-20	2007	R
10841	s8804	TV Show	Zombie Dumb	NaN	NaN	2019-07-01	2018	TV-Y7
10842	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	2019-11-01	2009	R
10843	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	2020-01-11	2006	PG

show_id	type	title	director	cast	date_added	release_year	rating
10844	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	2019-03-02	2015 TV-14

10845 rows × 11 columns

In [51]:

```
directorsplit=df['director'].apply(lambda x: str(x).split(',')).tolist()
directorsplits=pd.DataFrame(directorsplit,index=df['title'])
directorsplits=directorsplits.stack().reset_index()
directorsplits.drop('level_1',axis=1,inplace=True)
directorsplits.rename({0:'director_indv'},axis=1,inplace=True)
```

In [52]:

```
new=new.merge(directorsplits, left_on='title', right_on='title', how='left')
```

In [53]:

```
new.drop(['director'],axis=1,inplace=True)
```

In [54]:

new

Out[54]:

	show_id	type	title	cast	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Nan	2021-09-25	2020	PG-13	90 min Doc
1	s2	TV Show	Blood & Water	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021	TV-MA	2 Seasons Ir TV C
2	s3	TV Show	Ganglands	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	2021-09-24	2021	TV-MA	1 Season Ir TV
3	s4	TV Show	Jailbirds New Orleans	Nan	2021-09-24	2021	TV-MA	1 Season C
4	s5	TV Show	Kota Factory	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	2021-09-24	2021	TV-MA	2 Seasons Ir Re Sh
...	...	...	...	...	...	...	...	...
11907	s8803	Movie	Zodiac	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	2019-11-20	2007	R	158 min C
11908	s8804	TV Show	Zombie Dumb	Nan	2019-07-01	2018	TV-Y7	2 Seasons
11909	s8805	Movie	Zombieland	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	2019-11-01	2009	R	88 min HoI
11910	s8806	Movie	Zoom	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	2020-01-11	2006	PG	88 min Farr

	show_id	type	title	cast	date_added	release_year	rating	duration	
11911	s8807	Movie	Zubaan	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	2019-03-02	2015	TV-14	111 min	Ir Mo'

11912 rows × 11 columns



In [55]:

```
ind=new.loc[new['director_indv']=='nan'].index  
new_1=new.copy()  
new_1.drop(ind, axis=0, inplace=True)
```

In [56]:

new\_1

Out[56]:

	show_id	type	title	cast	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Nan	2021-09-25	2020	PG-13	90 min
2	s3	TV Show	Ganglands	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	2021-09-24	2021	TV-MA	1 Season
5	s6	TV Show	Midnight Mass	Kate Siegel, Zach Gilford, Hamish Linklater, H...	2021-09-24	2021	TV-MA	1 Season
6	s7	Movie	My Little Pony: A New Generation	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	2021-09-24	2021	PG	91 min
7	s7	Movie	My Little Pony: A New Generation	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	2021-09-24	2021	PG	91 min
...	...	...	...	...	...	...	...	...
11906	s8802	Movie	Zinzana	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	2016-03-09	2015	TV-MA	96 min
11907	s8803	Movie	Zodiac	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	2019-11-20	2007	R	158 min
11909	s8805	Movie	Zombieland	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	2019-11-01	2009	R	88 min

show_id	type	title	cast	date_added	release_year	rating	duration
11910	s8806	Movie	Zoom Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	2020-01-11	2006	PG	88 min
11911	s8807	Movie	Zubaan Vicky Kaushal, Sarah- Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14	111 min

8942 rows × 11 columns

In [57]:

```
new1=pd.crosstab(new_1['listed_coun'],new_1['director_indv'])
```

In [58]:

```
new1
```

Out[58]:

director_indv	A. L. Vijay	A. Raajdheep	A. Salaam	A. Murugadoss	A.R. Keluskar	Aadish	Aamir Bashir	Aamir Khan	Aanand Rai	Aaron Burns
listed_coun										
	0	0	0	0	0	0	0	0	0	0
<b>Afghanistan</b>	0	0	0	0	0	0	0	0	0	0
<b>Albania</b>	0	0	0	0	0	0	0	0	0	0
<b>Algeria</b>	0	0	0	0	0	0	0	0	0	0
<b>Angola</b>	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...
<b>Venezuela</b>	0	0	0	0	0	0	0	0	0	0
<b>Vietnam</b>	0	0	0	0	0	0	0	0	0	0
<b>West Germany</b>	0	0	0	0	0	0	0	0	0	0
<b>Zimbabwe</b>	0	0	0	0	0	0	0	0	0	0
<b>nan</b>	0	0	0	1	0	0	0	0	0	0

123 rows × 4993 columns

In [59]:

```
new2=new1.idxmax(axis=1).reset_index()
```

In [60]:

```
new2
```

Out[60]:

listed_coun	0
0	Najwa Najjar
1	Afghanistan Pieter-Jan De Pue
2	Albania Antonio Morabito
3	Algeria Maïwenn
4	Angola Chris Roland
...	...
118	Venezuela Aaron Woolf
119	Vietnam Bao Nhan
120	West Germany Christian Herrendoerfer
121	Zimbabwe Camilla Nielsson
122	nan Rajiv Chilaka

123 rows × 2 columns

In [61]:

```
new2.rename({0:'director'},axis=1,inplace=True)
```

In [62]:

```
new2.drop([0,122],inplace=True)
```

In [63]:

```
new2
```

```
# Below table shows the most liked director in each country
```

Out[63]:

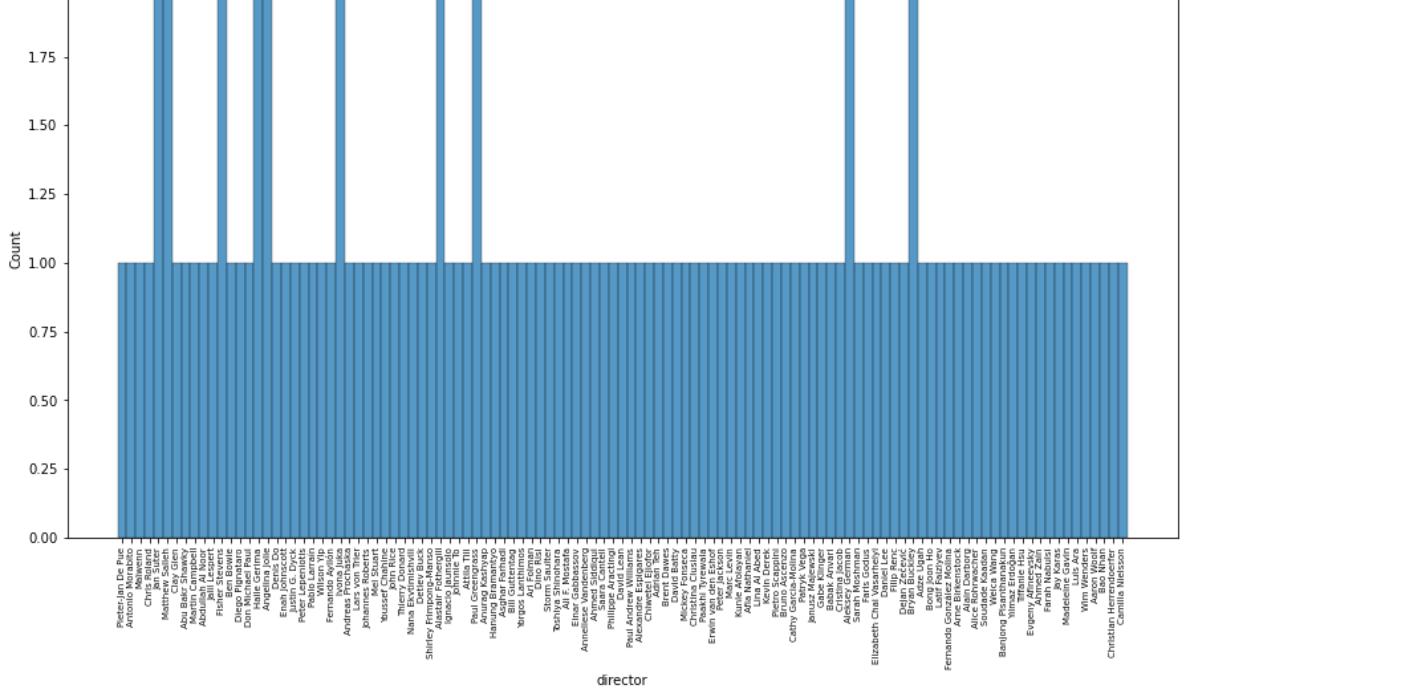
	listed_coun	director
1	Afghanistan	Pieter-Jan De Pue
2	Albania	Antonio Morabito
3	Algeria	Maïwenn
4	Angola	Chris Roland
5	Argentina	Jan Suter
...	...	...
117	Vatican City	Wim Wenders
118	Venezuela	Aaron Woolf
119	Vietnam	Bao Nhan
120	West Germany	Christian Herrendoerfer
121	Zimbabwe	Camilla Nielsson

121 rows × 2 columns

In [64]:



*e than 1 country. Jon Suter, Matthew Sullen, Fisher Stevens are the favorite directors in mo*



In [65]:

```
new_1['director_indv'].value_counts()
```

Out[65]:

Rajiv Chilaka	22
Jan Suter	21
Raúl Campos	19
Martin Scorsese	18
Steven Spielberg	18
..	
YC Tom Lee	1
Manu Ashokan	1
Anita Udeep	1
Alberto Arnaut Estrada	1
Mozez Singh	1

Name: director\_indv, Length: 4993, dtype: int64

In [66]:

```
li=(new_1['director_indv'].value_counts()>5).reset_index()
#sns.histplot(x=new_1['director_indv']>1)
#plt.xticks(rotation=90,fontsize=7)
#plt.gcf().set_size_inches(15, 8)
#plt.show()
```

In [67]:

```
li2=list(li.loc[li['director_indv']==True,'index'])
```

In [68]:

```
len(li2)
```

Out[68]:

166

In [69]:

```
new_22=new_1.loc[new_1['director_indv'].isin(li2)]
```

In [70]:

new\_22

Out[70]:

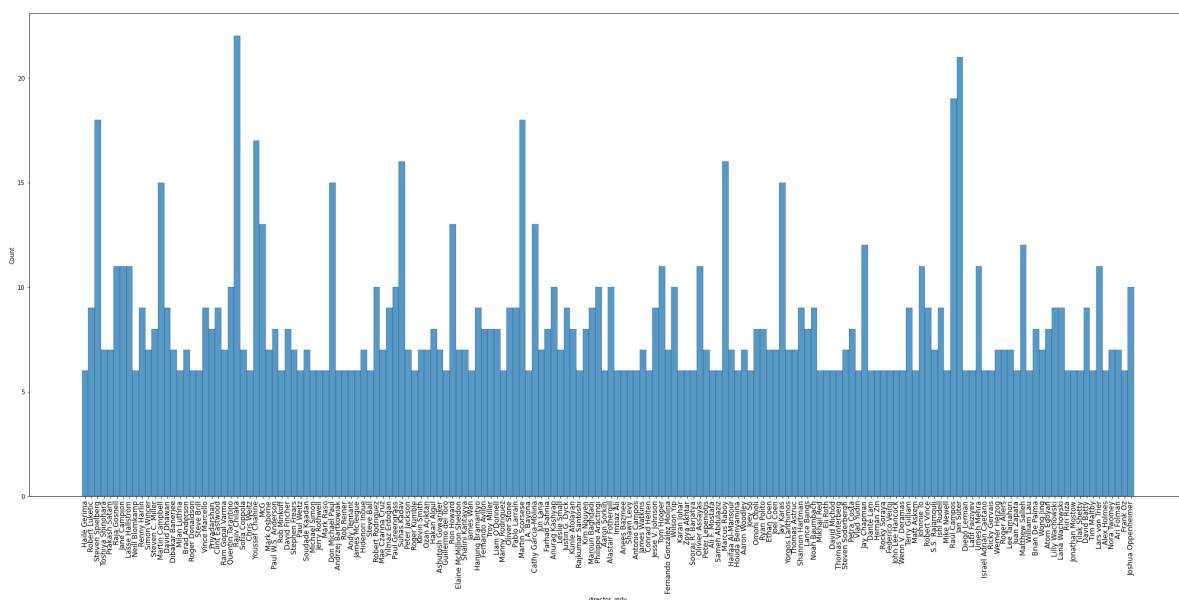
	show_id	type	title	cast	date_added	release_year	rating	duration	
8	s8	Movie	Sankofa	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	2021-09-24	1993	TV-MA	125 min	In...
9	s8	Movie	Sankofa	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	2021-09-24	1993	TV-MA	125 min	In...
10	s8	Movie	Sankofa	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	2021-09-24	1993	TV-MA	125 min	In...
11	s8	Movie	Sankofa	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	2021-09-24	1993	TV-MA	125 min	In...
12	s8	Movie	Sankofa	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	2021-09-24	1993	TV-MA	125 min	In...
...	...	...	...	...	...	...	...	...	...
11858	s8768	Movie	XXX: State of the Union	Ice Cube, Samuel L. Jackson, Willem Dafoe, Sco...	2019-01-01	2005	PG-13	101 min	In...
11877	s8783	Movie	Yoga Hosers	Harley Quinn Smith, Lily-Rose Depp, Johnny Dep...	2016-12-02	2016	PG-13	88 min	In...
11888	s8790	Movie	You Changed My Life	John Lloyd Cruz, Sarah Geronimo, Rayver Cruz, ...	2019-02-27	2009	TV-PG	116 min	In...
11892	s8794	Movie	Yours, Mine and Ours	Dennis Quaid, Rene Russo, Sean Faris, Katija P...	2019-11-20	2005	PG	88 min	In...

show_id	type	title	cast	date_added	release_year	rating	duration
11907	s8803	Movie	Zodiac	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	2019-11-20	2007	R 158 min

In [71]:

```
plot(x=new_22['director_indv'])  
.set_size_inches(35, 15)  
()
```

Shows the names of directors who have directed more than 5 movies. Count in the y axis shows that Chilaka, Youssef Chahine, McG Don Michael Paul, Suhas Kadav Ron Howard are the directors who

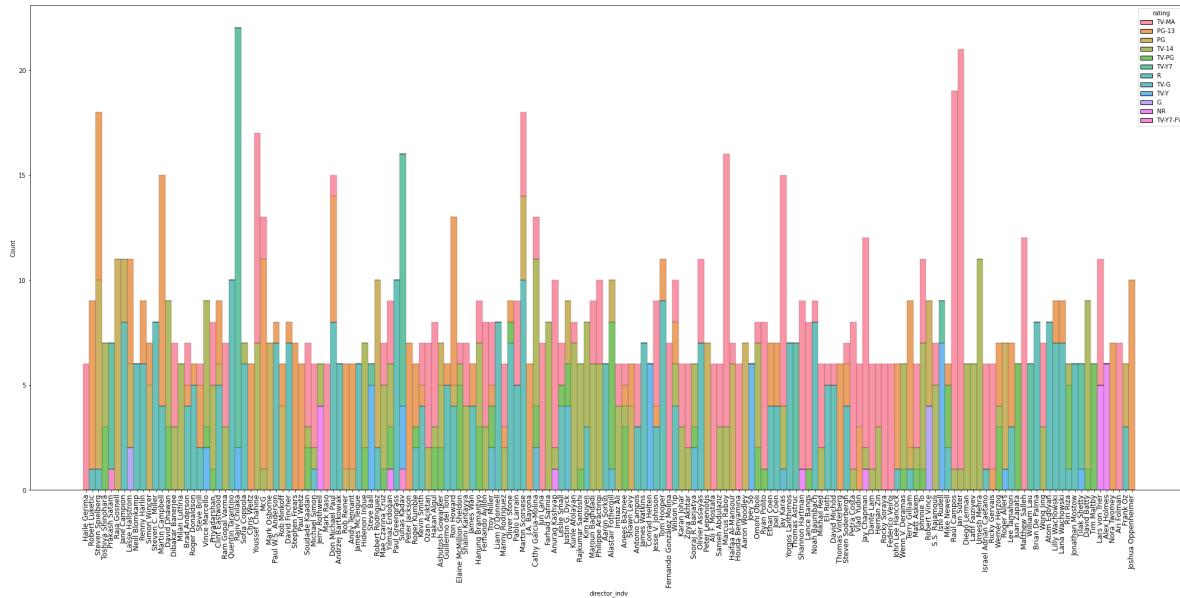


In [72]:

```

sns.histplot(x=new_22['director_indv'], hue=new_22['rating'], multiple="stack")
plt.xticks(rotation=90, fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()
#plot shows the rating movies directed by each director who has directed more than 5 movies

```

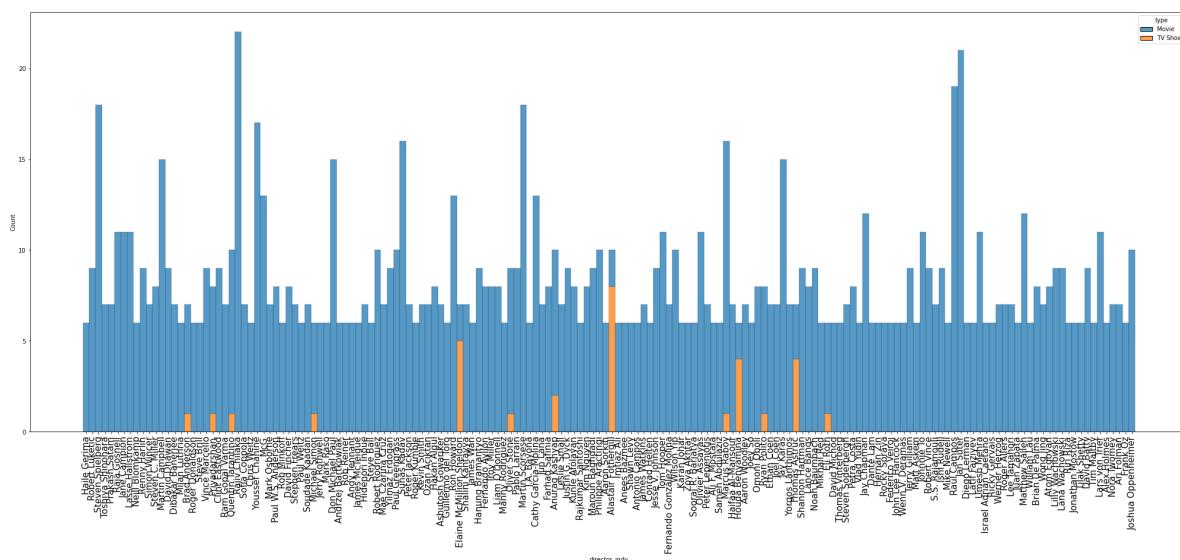


In [73]:

```

istplot(x=new_22['director_indv'], hue=new_22['type'], multiple="stack")
ticks(rotation=90, fontsize=15)
cf().set_size_inches(35, 13)
now()
# McMillion Sheldon, Alastair Fothergill, Houda Benyamina and Phomas Astruc are directors th

```

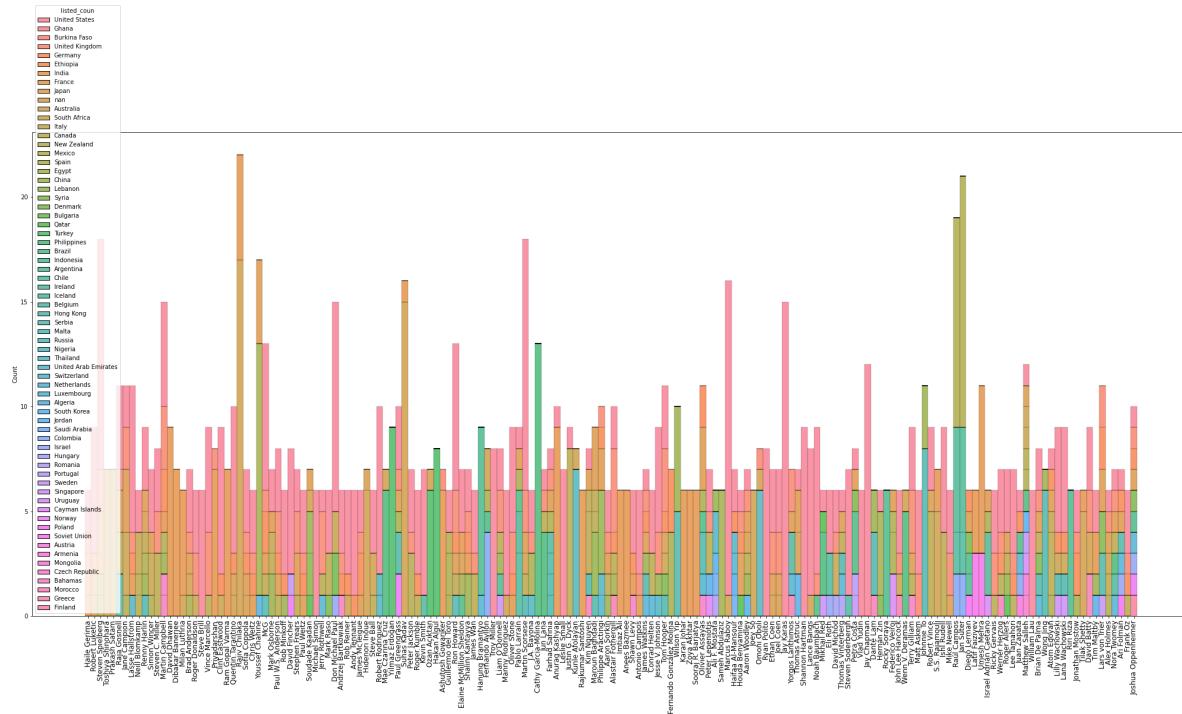


In [74]:

```

sns.histplot(x=new_22['director_indv'], hue=new_22['listed_coun'], multiple="stack")
plt.xticks(rotation=90, fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()
# Plot shows directors who has directed more than 5 movies that are released in each country

```



In [75]:

df

Out[75]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	2021-09-24	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	2021-09-24	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021-09-24	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	2019-11-20	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	2019-11-01	2009

show_id	type	title	director	cast	country	date_added	release_year	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	2020-01-11	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	2019-03-02	2015

8807 rows × 12 columns

In [76]:

```
listedsplit=df['listed_in'].apply(lambda x: str(x).split(', ')).tolist()
listedsplits=pd.DataFrame(listedsplit,index=df['title'])
listedsplits=listedsplits.stack().reset_index()
listedsplits.drop('level_1',axis=1,inplace=True)
listedsplits.rename({0:'listed_indv'},axis=1,inplace=True)
```

In [77]:

listedsplits

Out[77]:

	title	listed_indv
0	Dick Johnson Is Dead	Documentaries
1	Blood & Water	International TV Shows
2	Blood & Water	TV Dramas
3	Blood & Water	TV Mysteries
4	Ganglands	Crime TV Shows
...	...	...
19318	Zoom	Children & Family Movies
19319	Zoom	Comedies
19320	Zubaan	Dramas
19321	Zubaan	International Movies
19322	Zubaan	Music & Musicals

19323 rows × 2 columns

In [78]:

```
new=df.merge(listedsplits, left_on='title', right_on='title', how='left')
```

In [79]:

```
new.drop('listed_in',axis=1,inplace=True)
```

In [80]:

new

Out[80]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	P
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	T
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	T
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	T
4	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	2021-09-24	2021	T
...	...	...	...	...	...	...	...	...	...
19318	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	2020-01-11	2006	F
19319	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	2020-01-11	2006	F
19320	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	2019-03-02	2015	TV-

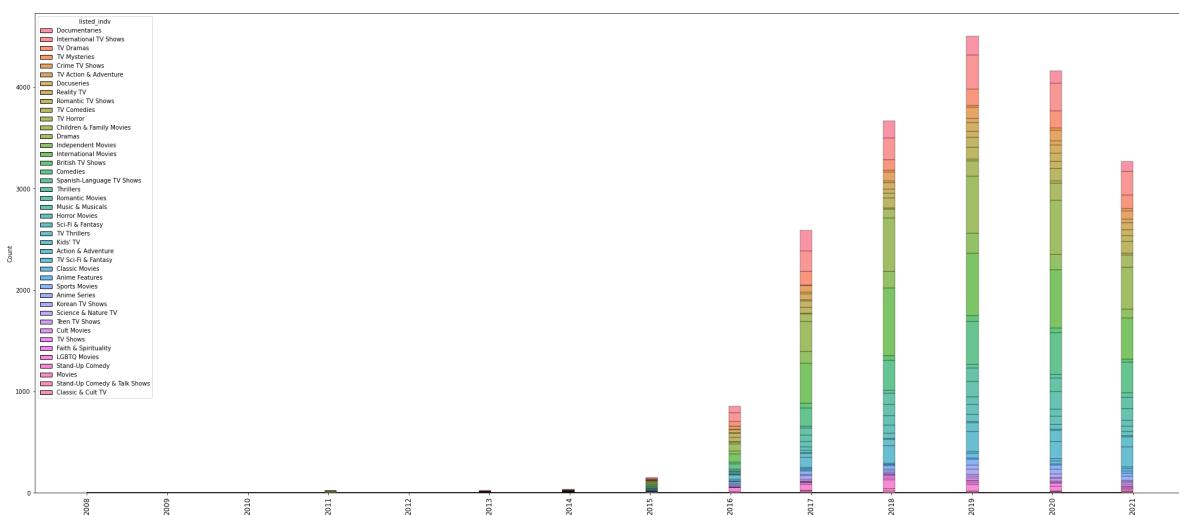
	show_id	type	title	director	cast	country	date_added	release_year	rating
19321	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	India	2019-03-02	2015	TV-
19322	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	India	2019-03-02	2015	TV-

19323 rows × 12 columns



In [81]:

```
sns.histplot(x=new['date_added'].dt.year,hue=new['listed_indv'],multiple="stack")
plt.xticks(np.arange(2008,2022),rotation=90,fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()
#plot shows how year wise the movie genres are Liked by people
# Every year people Like Dramas, Independent Movies and International movies more than other
```

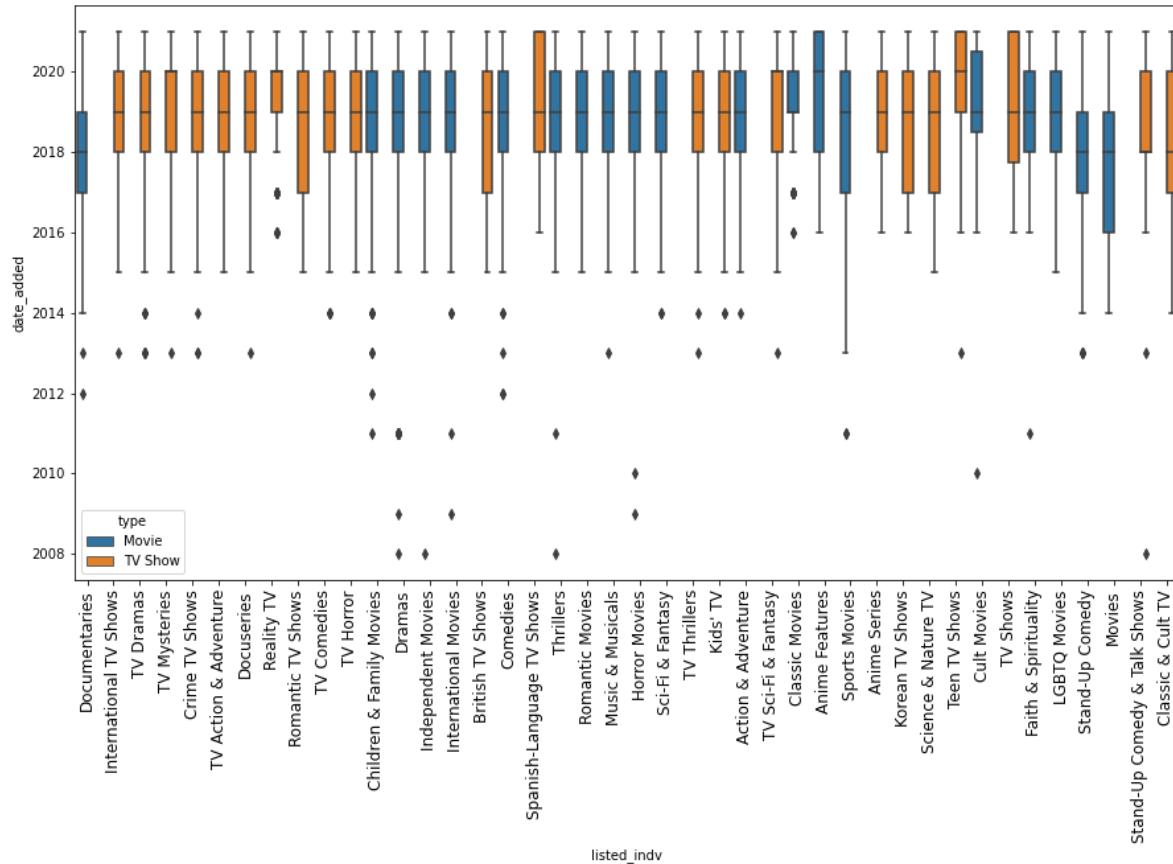


In [82]:

```

sns.boxplot(x=new['listed_in'],y=new['date_added'].dt.year,hue=new['type'])
plt.xticks(rotation=90,fontsize=12)
plt.gcf().set_size_inches(15, 8)
plt.show()
# Boxplot shows range of years movies were released with 'listed_in'

```



In [83]:

```
countrysplit=new['country'].apply(lambda x: str(x).split(', ')).tolist()
countrysplits=pd.DataFrame(countrysplit,index=new['title'])
countrysplits=countrysplits.stack().reset_index()
countrysplits.drop('level_1',axis=1,inplace=True)
countrysplits.rename({0:'listed_coun'},axis=1,inplace=True)
new=new.merge(countrysplits, left_on='title', right_on='title', how='left')
```

In [84]:

```
new.drop('country',axis=1,inplace=True)
```

In [85]:

```
new['listed_coun'].value_counts()>5
```

Out[85]:

```
United States      True
India             True
United Kingdom    True
nan               True
France            True
...
Somalia           False
Sudan              False
Nicaragua          False
Uganda             False
United States,     False
Name: listed_coun, Length: 128, dtype: bool
```

In [86]:

new

Out[86]:

	show_id	type	title	director	cast	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	2021-09-25	2020	PG-13	90
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021	TV-MA	Season
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021	TV-MA	Season
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021	TV-MA	Season
4	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021	TV-MA	Season
...	...	...	...	...	...	...	...	...	...
58563	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14	111
58564	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14	111
58565	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14	111

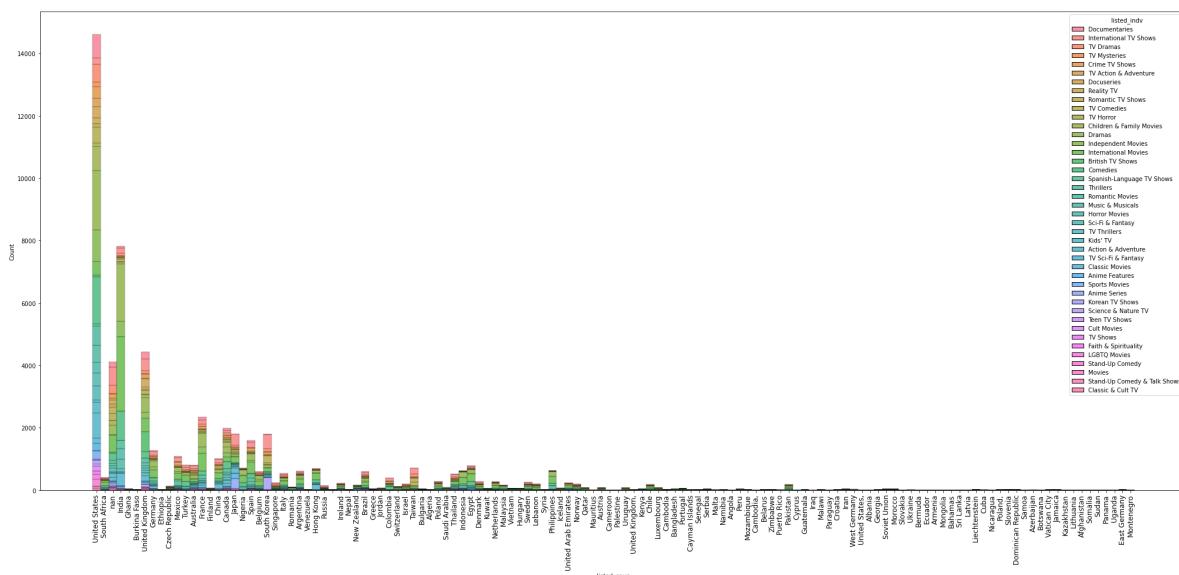
	show_id	type	title	director	cast	date_added	release_year	rating	duration
58566	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	2019-03-02	2015	TV-14	111
58567	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	2019-03-02	2015	TV-14	111

In [87]:

```

sns.histplot(x=new['listed_coun'], hue=new['listed_indv'], multiple="stack")
plt.xticks(rotation=90, fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()
# Plot shows country wise contour of 'listed_in'

```



In [88]:

```

li=(new['listed_coun'].value_counts()>1000).reset_index()
li2=list(li.loc[li['listed_coun']==True,'index'])
new_22=new.loc[new['listed_coun'].isin(li2)]

```

In [89]:

new\_22

Out[89]:

	show_id	type	title	director	cast	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	2021-09-25	2020	PG-13	90 m
10	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	2021-09-24	2021	TV-MA	Season
11	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	2021-09-24	2021	TV-MA	Season
12	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	2021-09-24	2021	TV-MA	Season
13	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	2021-09-24	2021	TV-MA	Season
...	...	...	...	...	...	...	...	...	...
58563	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14	111 m
58564	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14	111 m
58565	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14	111 m

show_id	type	title	director	cast	date_added	release_year	rating	duration
58566	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14 111 m
58567	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14 111 m

43845 rows × 12 columns

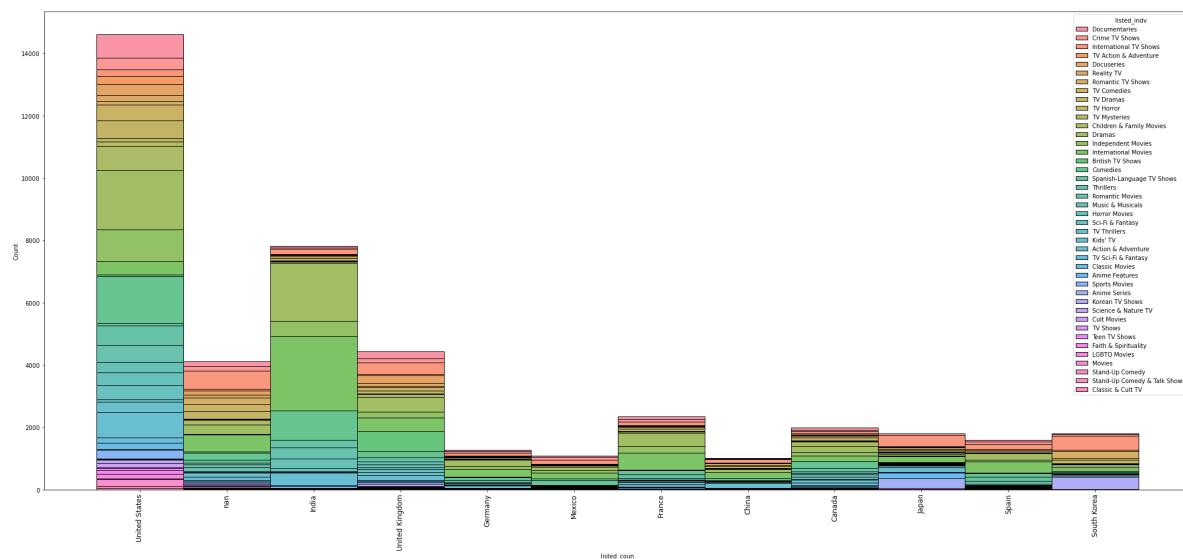


In [90]:

```

sns.histplot(x=new_22['listed_coun'], hue=new_22['listed_indv'], multiple="stack")
plt.xticks(rotation=90, fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()
# Plot shows the countries where more than 1000 movies are released and contour with 'liste

```

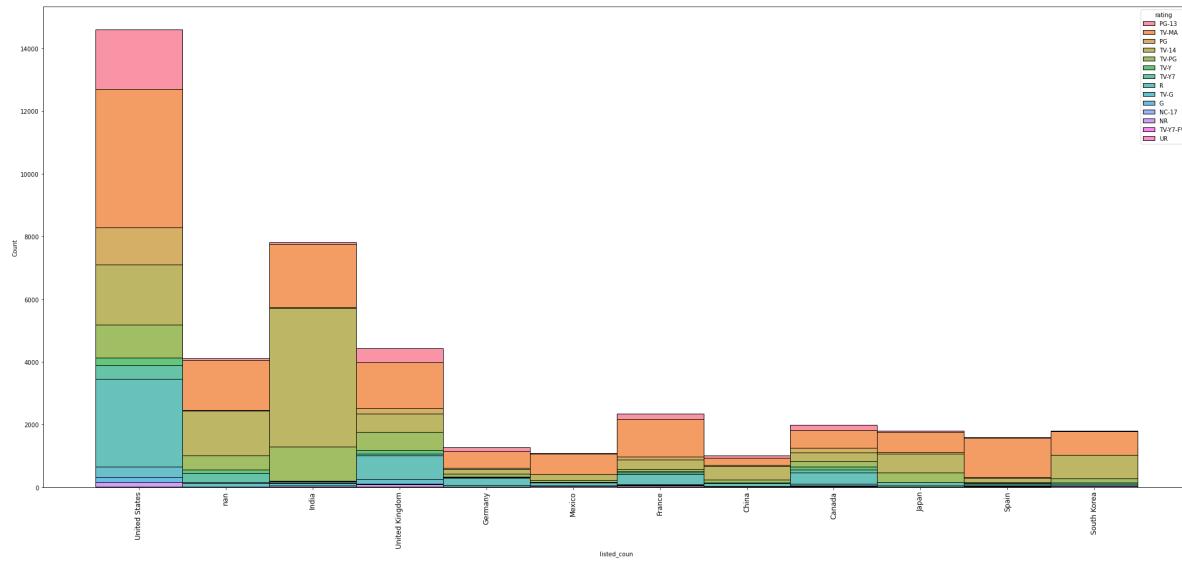


In [91]:

```

sns.histplot(x=new_22['listed_coun'], hue=new_22['rating'], multiple="stack")
plt.xticks(rotation=90, fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()
# Plot shows the countries where more than 1000 movies are released and contour with 'rating'

```



In [92]:

```

new1=pd.crosstab(new['listed_coun'],new['listed_indv'])
new2=new1.idxmax(axis=1).reset_index()
new2.rename({0:'listed_indv'},axis=1,inplace=True)
new2.drop([0,127],inplace=True)

```

In [93]:

```
new2  
# Listed_in category that people Like the most in each country
```

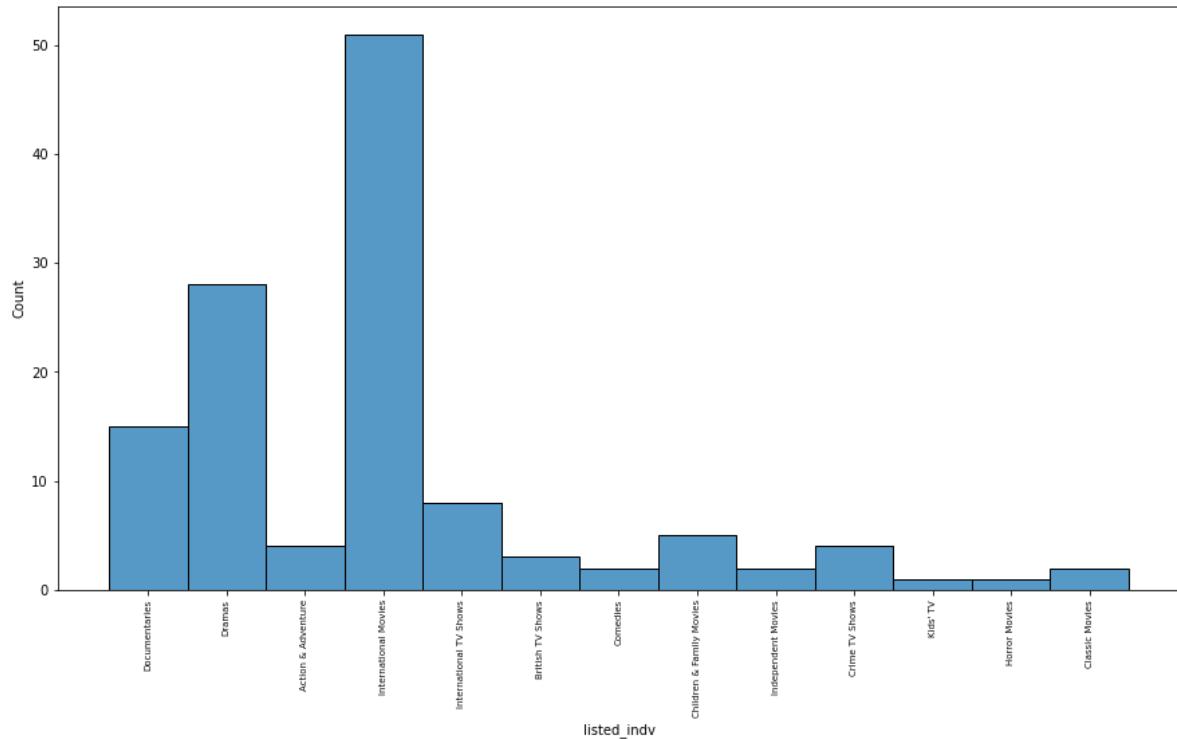
Out[93]:

	listed_coun	listed_indv
1	Afghanistan	Documentaries
2	Albania	Dramas
3	Algeria	Dramas
4	Angola	Action & Adventure
5	Argentina	International Movies
...	...	...
122	Vatican City	Documentaries
123	Venezuela	International Movies
124	Vietnam	International Movies
125	West Germany	International Movies
126	Zimbabwe	International Movies

126 rows × 2 columns

In [94]:

```
sns.histplot(x=new2['listed_indv'])
plt.xticks(rotation=90,fontsize=7)
plt.gcf().set_size_inches(15, 8)
plt.show()
# Plot shows 'Listed_in' vs count
# People like more of International movies, Documentries and Dramas
```



In [95]:

new

Out[95]:

	show_id	type	title	director	cast	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	2021-09-25	2020	PG-13	90
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021	TV-MA	Season
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021	TV-MA	Season
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021	TV-MA	Season
4	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	2021-09-24	2021	TV-MA	Season
...	...	...	...	...	...	...	...	...	...
58563	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14	111
58564	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14	111
58565	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	2019-03-02	2015	TV-14	111

	show_id	type	title	director	cast	date_added	release_year	rating	duration	
58566	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	2019-03-02	2015	TV-14	111	
58567	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	2019-03-02	2015	TV-14	111	

58568 rows × 12 columns



In [96]:

df

Out[96]:

	show_id	type	title	director	cast	country	date_added	release_year	rat
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	F
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	.
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	2021-09-24	2021	.
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	2021-09-24	2021	.
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021-09-24	2021	.
...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	2019-11-20	2007	.
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2018	TV-
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	2019-11-01	2009	.

show_id	type	title	director	cast	country	date_added	release_year	rat
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	2020-01-11	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	2019-03-02	2015 TV

8807 rows × 12 columns



In [97]:

```
new=df.copy()
castssplit=df['cast'].apply(lambda x: str(x).split(', ')).tolist()
castsplits=pd.DataFrame(castsplits, index=df['title'])
castsplits=castsplits.stack().reset_index()
castsplits.drop('level_1', axis=1, inplace=True)
castsplits.rename({0:'listed_cast'}, axis=1, inplace=True)
```

In [98]:

```
new=new.merge(castsplits, left_on='title', right_on='title', how='left')
```

In [99]:

new

Out[99]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	PG-13
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV-MA
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV-MA
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV-MA
4	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV-MA
...	...	...	...	...	...	...	...	...	...
64946	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	2019-03-02	2015	TV-14
64947	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	2019-03-02	2015	TV-14
64948	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	2019-03-02	2015	TV-14

	show_id	type	title	director	cast	country	date_added	release_year	rating
64949	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	2019-03-02	2015	TV-14
64950	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	2019-03-02	2015	TV-14

64951 rows × 13 columns



In [100]:

```
new['listed_cast'].value_counts()
```

Out[100]:

nan	825
Anupam Kher	43
Shah Rukh Khan	35
Julie Tejwani	33
Naseeruddin Shah	32
...	
Melanie Straub	1
Gabriela Maria Schmeide	1
Helena Zengel	1
Daniel Valenzuela	1
Chittaranjan Tripathy	1

Name: listed\_cast, Length: 36440, dtype: int64

In [101]:

```
li=(new['listed_cast'].value_counts()>15).reset_index()
li2=list(li.loc[li['listed_cast']==True,'index'])
new_22=new.loc[new['listed_cast'].isin(li2)]
```

In [ ]:

In [102]:

```
ind=new_22.loc[new_22['listed_cast']=='nan'].index
new_1=new_22.drop(ind, axis=0)
```

In [103]:

new\_1

Out[103]:

	show_id	type	title	director	cast	country	date_added	release_year
206	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	2021-09-20	2010
209	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	2021-09-20	2010
213	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	2021-09-20	2010
311	s40	TV Show	Chhota Bheem	NaN	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	India	2021-09-16	2021
312	s40	TV Show	Chhota Bheem	NaN	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	India	2021-09-16	2021
...	...	...	...	...	...	...	...	...
64661	s8771	Movie	Yaadein	Subhash Ghai	Jackie Shroff, Hrithik Roshan, Kareena Kapoor,...	India	2018-03-01	2001
64662	s8771	Movie	Yaadein	Subhash Ghai	Jackie Shroff, Hrithik Roshan, Kareena Kapoor,...	India	2018-03-01	2001
64685	s8773	Movie	Yamla Pagla Deewana 2	Sangeeth Sivan	Dharmendra, Sunny Deol, Bobby Deol, Neha Sharma...	India	2017-05-01	2013
64808	s8791	Movie	You Don't Mess with the Zohan	Dennis Dugan	Adam Sandler, John Turturro, Emmanuelle Chriqui...	United States	2019-09-01	2008

show_id	type	title	director	cast	country	date_added	release_year	
64886	s8799	Movie	Zed Plus	Chandra Prakash Dwivedi	Adil Hussain, Mona Singh, K.K. Raina, Sanjay M...	India	2019-12-31	2014

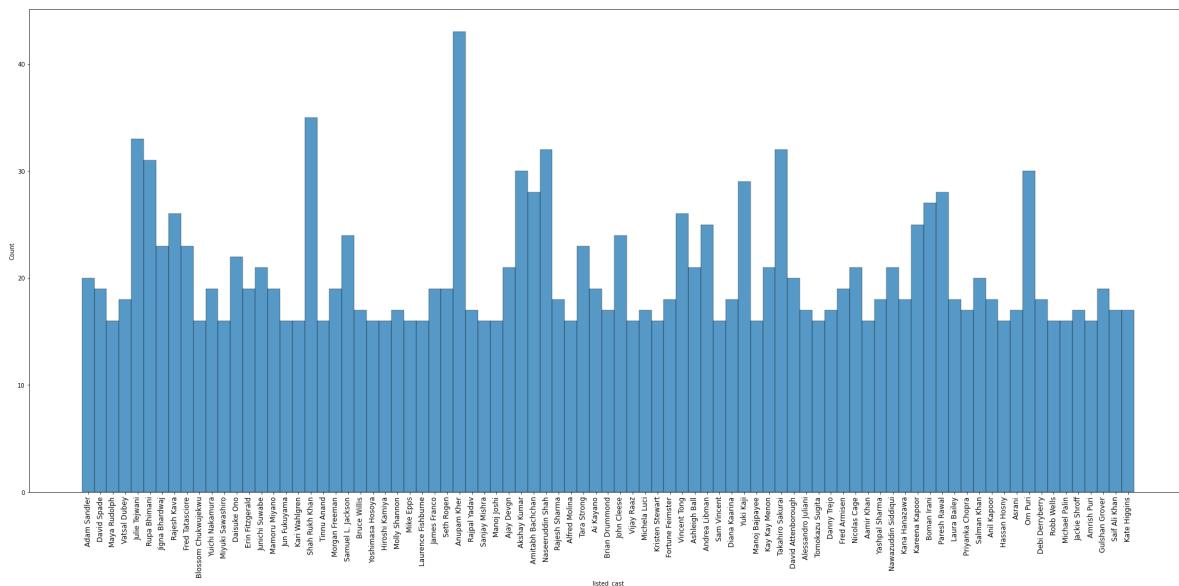
In [104]:

```

sns.histplot(x=new_1['listed_cast'])
plt.xticks(rotation=90, fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()

# Plot shows the actors who have acted in more than 15 movies

```



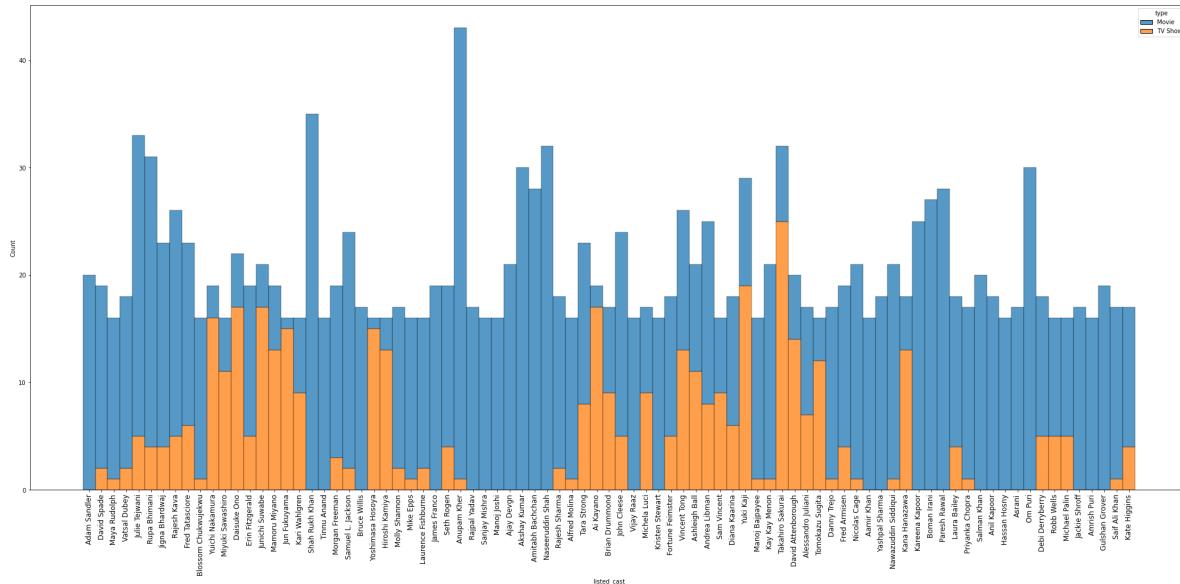
In [105]:

```

sns.histplot(x=new_1['listed_cast'], hue=new_1['type'], multiple='stack')
plt.xticks(rotation=90, fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()

# Plot shows the actors who have acted in more than 15 movies contour as type
# Miyuki Sawashiro, Junichi Suwabe, Jun Fukuyama, Ai Kayano acted more in TV shows than in

```



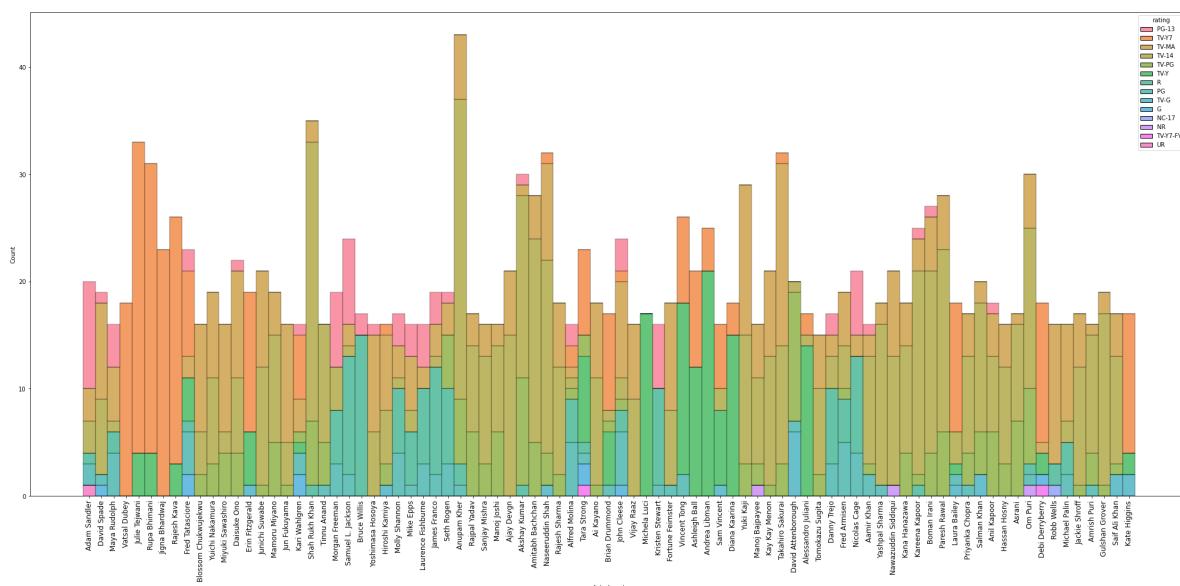
In [106]:

```

sns.histplot(x=new_1['listed_cast'], hue=new_1['rating'], multiple='stack')
plt.xticks(rotation=90, fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()

# Plot shows the actors vs number of different rating movies they have acted in

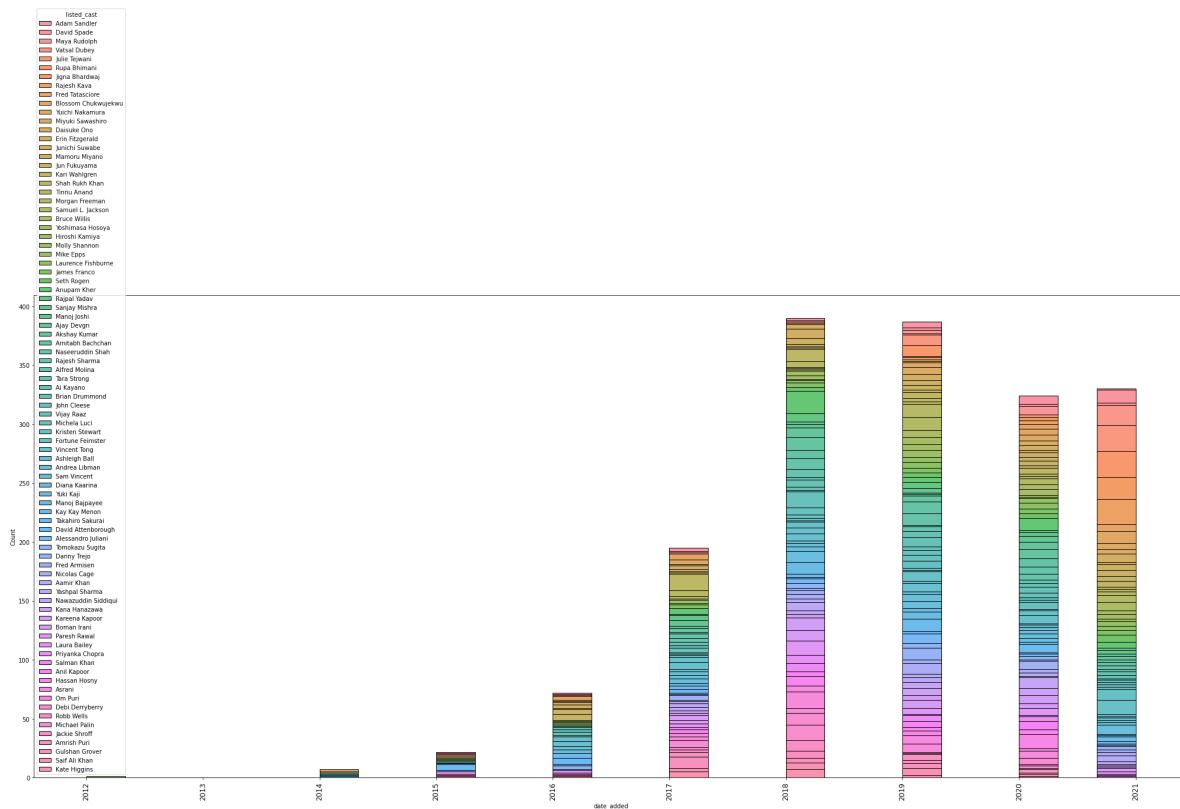
```



In [ ]:

In [107]:

```
sns.histplot(x=new_1['date_added'].dt.year,hue=new_1['listed_cast'],multiple='stack')
plt.xticks(np.arange(2012,2022,1),rotation=90,fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()
```



In [108]:

new\_1

Out[108]:

	show_id	type	title	director	cast	country	date_added	release_year
206	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	2021-09-20	2010
209	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	2021-09-20	2010
213	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	2021-09-20	2010
311	s40	TV Show	Chhota Bheem	NaN	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	India	2021-09-16	2021
312	s40	TV Show	Chhota Bheem	NaN	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	India	2021-09-16	2021
...	...	...	...	...	...	...	...	...
64661	s8771	Movie	Yaadein	Subhash Ghai	Jackie Shroff, Hrithik Roshan, Kareena Kapoor,...	India	2018-03-01	2001
64662	s8771	Movie	Yaadein	Subhash Ghai	Jackie Shroff, Hrithik Roshan, Kareena Kapoor,...	India	2018-03-01	2001
64685	s8773	Movie	Yamla Pagla Deewana 2	Sangeeth Sivan	Dharmendra, Sunny Deol, Bobby Deol, Neha Sharma...	India	2017-05-01	2013
64808	s8791	Movie	You Don't Mess with the Zohan	Dennis Dugan	Adam Sandler, John Turturro, Emmanuelle Chriqui...	United States	2019-09-01	2008

	show_id	type	title	director	cast	country	date_added	release_year	
64886	s8799	Movie	Zed Plus	Chandra Prakash Dwivedi	Adil Hussain, Mona Singh, K.K. Raina, Sanjay M...	India	2019-12-31	2014	

1729 rows × 13 columns

In [109]:

```
listedsplit=new_1['listed_in'].apply(lambda x: str(x).split(', ')).tolist()
listedsplits=pd.DataFrame(listedsplit,index=new_1['title'])
listedsplits=listedsplits.stack().reset_index()
listedsplits.drop('level_1',axis=1,inplace=True)
listedsplits.rename({0:'listed_inin'},axis=1,inplace=True)
new=new_1.merge(listedsplits,left_on='title',right_on='title',how='left')
```

In [110]:

new

Out[110]:

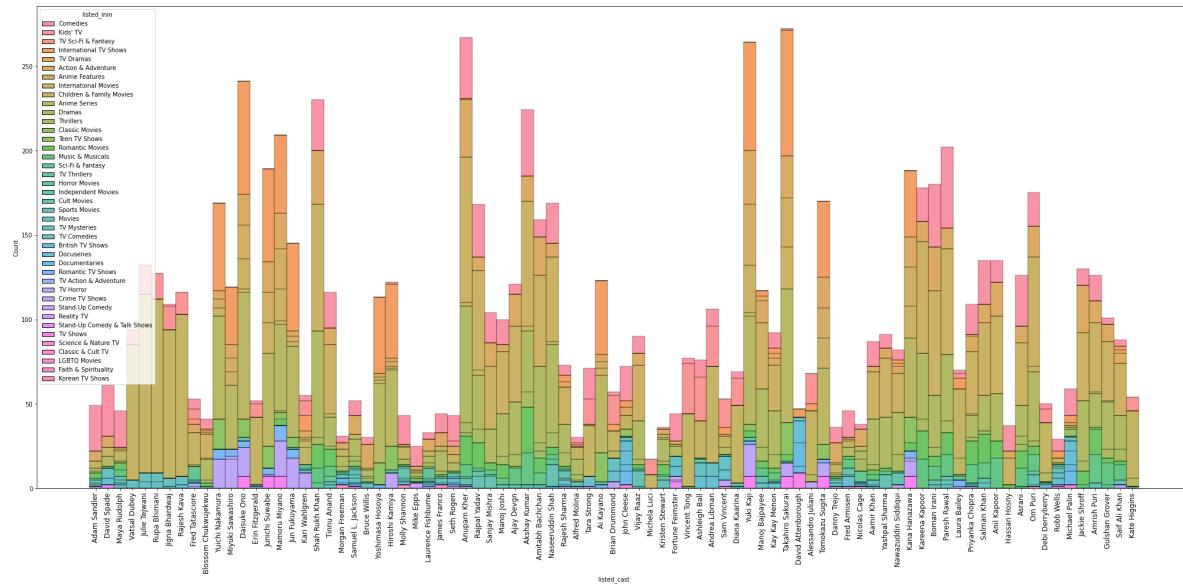
	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	2021-09-20	2010	PG-13
1	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	2021-09-20	2010	PG-13
2	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	2021-09-20	2010	PG-13
3	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	2021-09-20	2010	PG-13
4	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	2021-09-20	2010	PG-13
...	...	...	...	...	...	...	...	...	...
8772	s8791	Movie	You Don't Mess with the Zohan	Dennis Dugan	Adam Sandler, John Turturro, Emmanuelle Chriqui...	United States	2019-09-01	2008	UR
8773	s8791	Movie	You Don't Mess with the Zohan	Dennis Dugan	Adam Sandler, John Turturro, Emmanuelle Chriqui...	United States	2019-09-01	2008	UR
8774	s8799	Movie	Zed Plus	Chandra Prakash Dwivedi	Adil Hussain, Mona Singh, K.K. Raina, Sanjay M...	India	2019-12-31	2014	TV-MA
8775	s8799	Movie	Zed Plus	Chandra Prakash Dwivedi	Adil Hussain, Mona Singh, K.K. Raina, Sanjay M...	India	2019-12-31	2014	TV-MA

show_id	type	title	director	cast	country	date_added	release_year	rating	
8776	s8799	Movie	Zed Plus	Chandra Prakash Dwivedi	Adil Hussain, Mona Singh, K.K. Raina, Sanjay M...	India	2019-12-31	2014	TV-MA

8777 rows × 14 columns

In [111]:

```
sns.histplot(x=new['listed_cast'],hue=new['listed_inin'],multiple='stack')
plt.xticks(rotation=90,fontsize=12)
plt.gcf().set_size_inches(35, 15)
plt.show()
# Plot shows the actors vs number of different 'listed_in' thay have acted in
```



In [112]:

df

Out[112]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	2021-09-24	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	2021-09-24	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021-09-24	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	2019-11-20	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	2019-07-01	2018
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	2019-11-01	2009

show_id	type	title	director	cast	country	date_added	release_year	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	2020-01-11	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	2019-03-02	2015

8807 rows × 12 columns

In [113]:

```

new=df.copy()
consplit=df['country'].apply(lambda x: str(x).split(', ')).tolist()
consplits=pd.DataFrame(consplit,index=df['title'])
consplits=consplits.stack().reset_index()
consplits.drop('level_1',axis=1,inplace=True)
consplits.rename({0:'listed_con'},axis=1,inplace=True)
new=new.merge(consplits,left_on='title',right_on='title',how='left')

castssplit=df['cast'].apply(lambda x: str(x).split(', ')).tolist()
castsplits=pd.DataFrame(castsplits, index=df['title'])
castsplits=castsplits.stack().reset_index()
castsplits.drop('level_1',axis=1,inplace=True)
castsplits.rename({0:'listed_cast'},axis=1,inplace=True)
new=new.merge(castsplits, left_on='title', right_on='title', how='left')

```

In [114]:

new

Out[114]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2020	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	
4	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	
...	...	...	...	...	...	...	...	...	...
81703	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	2019-03-02	2015	TV
81704	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	2019-03-02	2015	TV
81705	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	2019-03-02	2015	TV

show_id	type	title	director	cast	country	date_added	release_year	rating
81706	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	2019-03-02	2015 TV
81707	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	2019-03-02	2015 TV

In [115]:

```
ind=new.loc[new['listed_cast']=='nan'].index
new=new.drop(ind,axis=0)
new1=pd.crosstab(new['listed_con'],new['listed_cast'])
new2=new1.idxmax(axis=1).reset_index()
new2.rename({0:'listed_indv'},axis=1,inplace=True)
#new2.drop([0,127],inplace=True)
```

In [116]:

```
new2
# Table below shows favorite actor in each country
```

Out[116]:

	listed_con	listed_indv
0		Areen Omari
1	Afghanistan	Sohrab Nazari
2	Albania	Claudio Santamaria
3	Algeria	Abdalla Mahmoud
4	Angola	Cigano Satyohamba
...	...	...
115	Venezuela	Andrea Garrote
116	Vietnam	Mai Cat Vi
117	West Germany	Christopher Lee
118	Zimbabwe	Charmaine Mujeri
119	nan	Julie Tejwani

120 rows × 2 columns

In [117]:

```
ind=new2.loc[new2['listed_indv']=='nan'].index
new_1=new2.drop(ind,axis=0)
```

In [118]:

```
new_1
```

Out[118]:

	listed_con	listed_indv
0		Areen Omari
1	Afghanistan	Sohrab Nazari
2	Albania	Claudio Santamaria
3	Algeria	Abdalla Mahmoud
4	Angola	Cigano Satyohamba
...	...	...
115	Venezuela	Andrea Garrote
116	Vietnam	Mai Cat Vi
117	West Germany	Christopher Lee
118	Zimbabwe	Charmaine Mujeri
119	nan	Julie Tejwani

120 rows × 2 columns

In [119]:

```
listedsplit=new['listed_in'].apply(lambda x: str(x).split(', ')).tolist()
listedsplits=pd.DataFrame(listedsplit,index=new['title'])
listedsplits=listedsplits.stack().reset_index()
listedsplits.drop('level_1',axis=1,inplace=True)
listedsplits.rename({0:'listed_inin'},axis=1,inplace=True)
new_11=new.merge(listedsplits, left_on='title', right_on='title', how='left')
```

In [120]:

new\_11

Out[120]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV M.
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV M.
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV M.
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV M.
4	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV M.
...	...	...	...	...	...	...	...	...	...
3108419	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	2019-03-02	2015	TV-1
3108420	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	2019-03-02	2015	TV-1
3108421	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chan...	India	2019-03-02	2015	TV-1

	show_id	type	title	director	cast	country	date_added	release_year	rating
3108422	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	India	2019-03-02	2015	TV-1
3108423	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	India	2019-03-02	2015	TV-1

3108424 rows × 15 columns

In [121]:

```
ind=new_11.loc[new_11['listed_inin']=='nan'].index
new_11=new_11.drop(ind,axis=0)
ind=new_11.loc[new_11['listed_cast']=='nan'].index
new_11=new_11.drop(ind,axis=0)
new1=pd.crosstab(new_11['listed_inin'],new_11['listed_cast'])
new2=new1.idxmax(axis=1).reset_index()
new2.rename({0:'listed_indv'},axis=1,inplace=True)
#new2.drop([0,127],inplace=True)
```

In [122]:

new2

# Table below shows the most number of times a actor has acted in a particular 'listed\_in'

Out[122]:

	listed_inin	listed_indv
0	Action & Adventure	Kelly Reilly
1	Anime Features	Yuki Kaji
2	Anime Series	Takahiro Sakurai
3	British TV Shows	Kate Harbour
4	Children & Family Movies	James Franco
5	Classic & Cult TV	Don S. Davis
6	Classic Movies	Harvey Keitel
7	Comedies	John Cleese
8	Crime TV Shows	Raúl Méndez
9	Cult Movies	Alison Pill
10	Documentaries	Ahmed Hassan
11	Docuseries	David Attenborough
12	Dramas	Eddie Marsan
13	Faith & Spirituality	Abdelilah Wahbi
14	Horror Movies	Wentworth Miller
15	Independent Movies	Jennifer Ehle
16	International Movies	Matthias Schoenaerts
17	International TV Shows	Lee Ingleby
18	Kids' TV	Vincent Tong
19	Korean TV Shows	Seo Hyun-jin
20	LGBTQ Movies	Allan Paule
21	Movies	Ben Diskin
22	Music & Musicals	Zac Efron
23	Reality TV	Anderson Silva
24	Romantic Movies	Ben Whishaw
25	Romantic TV Shows	Alfred Enoch
26	Sci-Fi & Fantasy	Paul Giamatti
27	Science & Nature TV	David Attenborough
28	Spanish-Language TV Shows	Carmen Villalobos
29	Sports Movies	Jay Baruchel
30	Stand-Up Comedy	Jack Whitehall
31	Stand-Up Comedy & Talk Shows	Nazeem Hussain
32	TV Action & Adventure	José María Yazpik

	listed_inin	listed_indv
33	TV Comedies	John Sparkes
34	TV Dramas	Christopher Lee
35	TV Horror	Jon Jon Briones
36	TV Mysteries	Danica Curcic
37	TV Sci-Fi & Fantasy	Lena Headey
38	TV Shows	Aditi Balan
39	TV Thrillers	Acushla-Tara Kupe
40	Teen TV Shows	Yuichi Nakamura
41	Thrillers	Tom Wilkinson

In [123]:

```
df1=df.loc[df['duration'].str.contains('min')]
df2=df.loc[df['duration'].str.contains('Sea')]
```

In [124]:

```
l1df1=[ ]
```

In [ ]:

In [125]:

```
for i in df1['duration']:
    ans=int(i.split(' ')[0])
    l1df1.append(ans)
```

In [126]:

```
df1['dur']=l1df1
```

C:\Users\hp\AppData\Local\Temp\ipykernel\_16588\1913062087.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy))

```
df1['dur']=l1df1
```

In [127]:

df1

Out[127]:

	show_id	type	title	director	cast	country	date_added	release_y
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	2021-09-25	2
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	NaN	2021-09-24	2
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	2021-09-24	1
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	2021-09-24	2
12	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	2021-09-23	2
...	...	...	...	...	...	...	...	...
8801	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	2016-03-09	2
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey Jr...	United States	2019-11-20	2
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	2019-11-01	2
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	2020-01-11	2

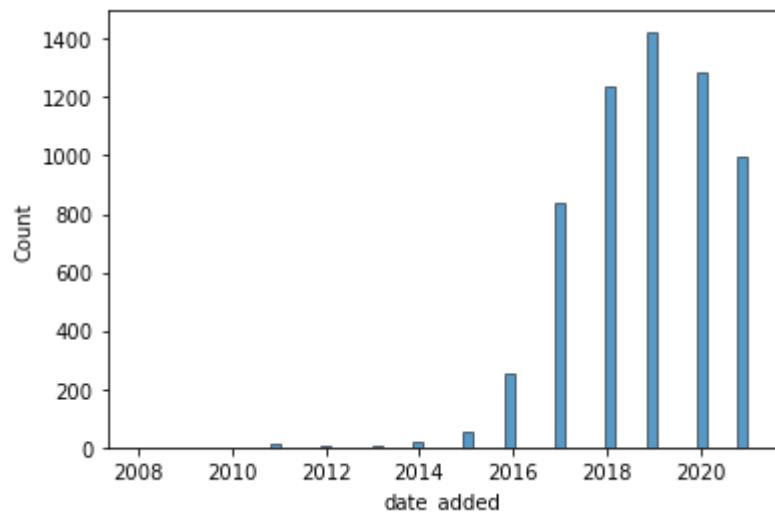
show_id	type	title	director	cast	country	date_added	release_y	
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanani...	India	2019-03-02	2

In [128]:

```
sns.histplot(x=df1['date_added'].dt.year)
#plot shows the trend in movies with respect to years
```

Out[128]:

&lt;AxesSubplot:xlabel='date\_added', ylabel='Count'&gt;

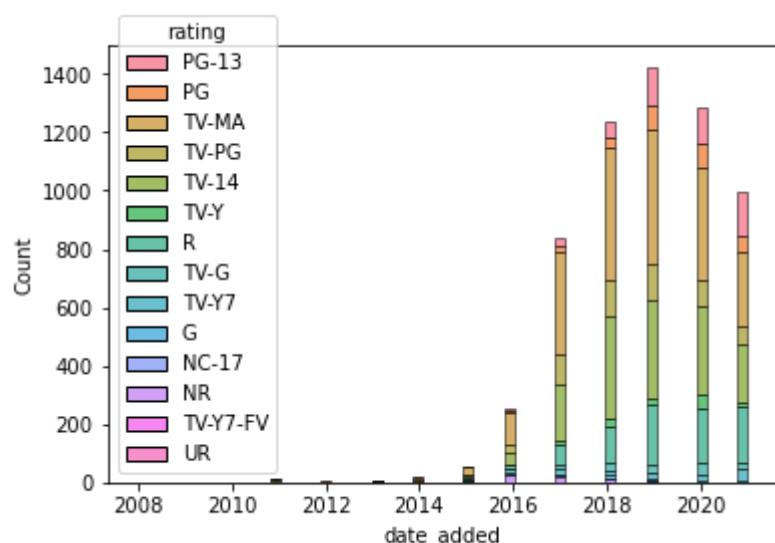


In [132]:

```
sns.histplot(x=df1['date_added'].dt.year,hue=df1['rating'],multiple='stack')
```

Out[132]:

&lt;AxesSubplot:xlabel='date\_added', ylabel='Count'&gt;

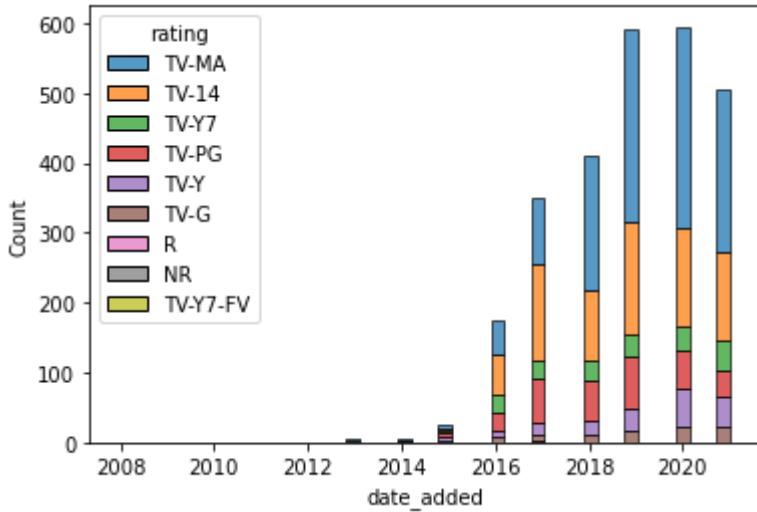


In [133]:

```
sns.histplot(x=df2['date_added'].dt.year,hue=df2['rating'],multiple='stack')
```

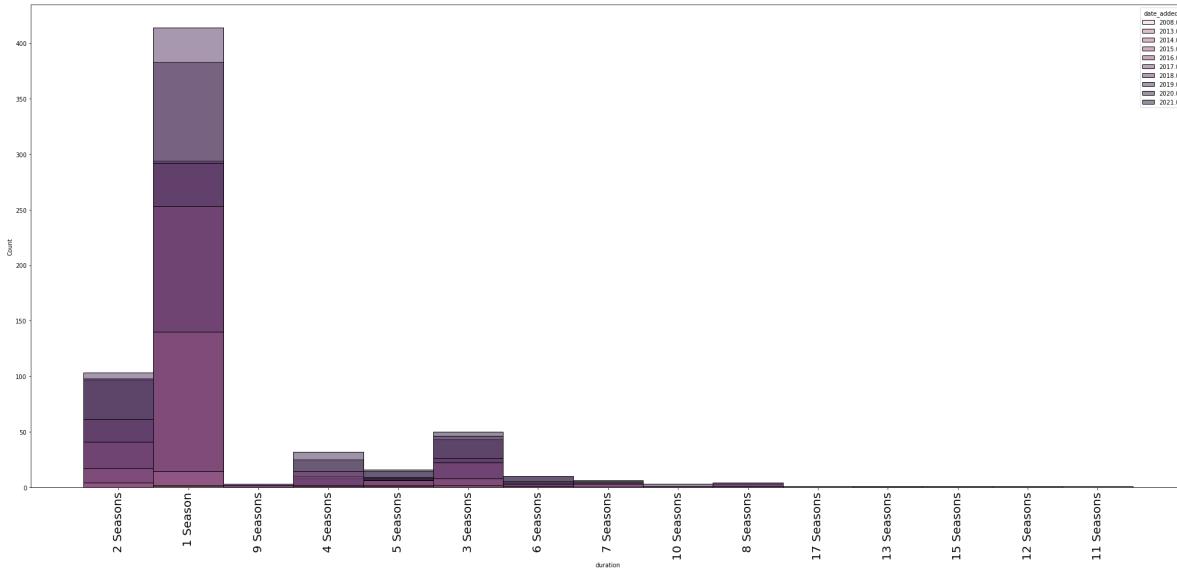
Out[133]:

<AxesSubplot:xlabel='date\_added', ylabel='Count'>



In [134]:

```
sns.histplot(x=df2['duration'],hue=df2['date_added'].dt.year)
plt.xticks(rotation=90,fontsize=20)
plt.gcf().set_size_inches(35, 15)
plt.show()
```



In [139]:

```
df1.describe()  
# Mean duration of the movies are 99 min
```

Out[139]:

	release_year	dur
<b>count</b>	6131.000000	6131.000000
<b>mean</b>	2013.121514	99.564998
<b>std</b>	9.678169	28.289504
<b>min</b>	1942.000000	3.000000
<b>25%</b>	2012.000000	87.000000
<b>50%</b>	2016.000000	98.000000
<b>75%</b>	2018.000000	114.000000
<b>max</b>	2021.000000	312.000000

In [ ]:

In [ ]: