

In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:

```
df=pd.read_csv('netflix_titles.csv')
```

In [3]:

```
df.head(10)
```

3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV
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 ...
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
			My Little	Robert	Vanessa Hudgens,						

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   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 [5]:

```
df.shape
```

Out[5]:

(8807, 12)

In [6]:

```
df.describe()
```

Out[6]:

	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 [7]:

```
df.describe(include='object').T
```

Out[7]:

	count	unique	top	freq
show_id	8807	8807	s3460	1
type	8807	2	Movie	6131
title	8807	8807	Zig & Sharko	1
director	6173	4528	Rajiv Chilaka	19
cast	7982	7692	David Attenborough	19
country	7976	748	United States	2818
date_added	8797	1767	January 1, 2020	109
rating	8803	17	TV-MA	3207
duration	8804	220	1 Season	1793
listed_in	8807	514	Dramas, International Movies	362
description	8807	8775	Paranormal activity at a lush, abandoned prope...	4

In [8]:

```
df.isnull().sum()/len(df)*100
```

Out[8]:

```
show_id      0.000000
type         0.000000
title        0.000000
director     29.908028
cast         9.367549
country      9.435676
date_added   0.113546
release_year  0.000000
rating       0.045418
duration     0.034064
listed_in    0.000000
description   0.000000
dtype: float64
```

Movies vs TV Show Comparison

In [9]:

```
df['type'].value_counts(normalize=True)*100
```

Out[9]:

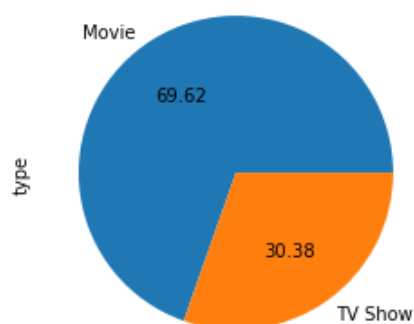
```
Movie      69.615079
TV Show    30.384921
Name: type, dtype: float64
```

In [10]:

```
df['type'].value_counts().plot(kind='pie', autopct="%.2f")
```

Out[10]:

<AxesSubplot:ylabel='type'>



In [11]:

```
df.head()
```

Out[11]:

	show_id	type		title	director	cast	country	date_added	release_year	rating	duration	
0	s1	Movie		Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Docu
1	s2	TV Show		Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	Int TV S Dr
2	s3	TV Show		Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Int TV S
3	s4	TV Show		Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Dr f
4	s5	TV Show		Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	Int T Roi Shc

In [12]:

```
df['cast'].apply(lambda x: str(x).split(', ')).tolist()
```

Out[12]:

```
[['nan'],
 ['Ama Qamata',
  'Khosi Ngema',
  'Gail Mabalane',
  'Thabang Molaba',
  'Dillon Windvogel',
  'Natasha Thahane',
  'Arno Greeff',
  'Xolile Tshabalala',
  'Getmore Sithole',
  'Cindy Mahlangu',
  'Ryle De Morny',
  'Greteli Fincham',
  'Sello Maake Ka-Ncube',
  'Odwa Gwanya',
  'Mekaila Mathys',
  'Sandi Schultz',
  'Duane Williams']
```

Splitting Cast, Genres, Directors

In [13]:

```
constraint_cast=df['cast'].apply(lambda x: str(x).split(',')).tolist()
constraint_director=df['director'].apply(lambda x: str(x).split(',')).tolist()
constraint_genre = df['listed_in'].apply(lambda x: str(x).split(',')).tolist()

df_new_cast=pd.DataFrame(constraint_cast,index=df['title']).stack()
df_new_cast = pd.DataFrame(df_new_cast)
df_new_cast.reset_index(inplace=True) #to give index to a both title and cast, otherwise will show error
df_new_cast = df_new_cast[['title',0]]
df_new_cast.columns = ['title','cast']

df_new_director=pd.DataFrame(constraint_director,index=df['title']).stack()
df_new_director = pd.DataFrame(df_new_director)
df_new_director.reset_index(inplace=True) #to give index to a both title and cast, otherwise will show error
df_new_director = df_new_director[['title',0]]
df_new_director.columns = ['title','director']

df_new_genre=pd.DataFrame(constraint_genre,index=df['title']).stack()
df_new_genre = pd.DataFrame(df_new_genre)
df_new_genre.reset_index(inplace=True) #to give index to a both title and cast, otherwise will show error
df_new_genre = df_new_genre[['title',0]]
df_new_genre.columns = ['title','genre']

df_new_cast_no_nan = df_new_cast.replace('nan',np.nan).dropna()
df_new_director_no_nan = df_new_director.replace('nan',np.nan).dropna()
df_new_genre_no_nan = df_new_genre.replace('nan',np.nan).dropna()

df_new_cast_no_nan["cast"].value_counts().head(1), df_new_director_no_nan["director"].value_counts().head(1)
```

Out[13]:

```
(Anupam Kher      43
 Name: cast, dtype: int64,
 Rajiv Chilaka   22
 Name: director, dtype: int64,
 International Movies  2752
 Name: genre, dtype: int64)
```

Basic Metrics

Most popular in each country and year:

1. Cast

In [14]:

```
df_cast=df.reset_index().merge(df_new_cast_no_nan, on='title', how='inner')
df_cast.groupby(["country","cast_y"]['title'].nunique().sort_values(ascending=False).head(1)
```

Out[14]:

```
country  cast_y
India    Anupam Kher    40
Name: title, dtype: int64
```

2. Director

In [15]:

```
df_director=df.reset_index().merge(df_new_director_no_nan, on='title', how='inner')
df_director.groupby(["country", "director_y"])['title'].nunique().sort_values(ascending=False).head(1)
```

Out[15]:

```
country    director_y
United States  Marcus Raboy    15
Name: title, dtype: int64
```

3. Genre

In [16]:

```
df_genre=df.reset_index().merge(df_new_genre_no_nan, on='title', how='inner')
df_genre.groupby(["country", "genre"])['title'].nunique().sort_values(ascending=False).head(1)
```

Out[16]:

```
country  genre
India    International Movies    817
Name: title, dtype: int64
```

Fastest addition after release

In [17]:

```
# Directors whose movies were added within 1 years of release

import seaborn as sns
import matplotlib.pyplot as plt

df["date_added_period"] = pd.to_datetime(df["date_added"]).dt.year - df["release_year"]
df_director=df.reset_index().merge(df_new_director_no_nan, on='title', how='inner')
df_director_added_period_max = df_director.groupby("director_y").filter(lambda x:x["date_added_period"]
```

In [18]:

```
df_director_added_period = df.reset_index().merge(df_director_added_period_max, on="title", how='inner')
df_director_added_period.groupby(["release_year_x", "director_y"])['title'].nunique().sort_values(ascending=False).head(1)
```

Out[18]:

```
release_year_x  director_y
2018            Raúl Campos    12
                Jan Suter     12
2021            Rathindran R Prasad    5
2017            Jan Suter     5
2020            Stanley Moore    4
                ..
2017            Sonia Lowman     1
                Stefan Brogren    1
                Stephen Cone     1
                Steven Brill     1
2021            Zhao Ji        1
Name: title, Length: 2824, dtype: int64
```

In [19]:

```
# countries with movie content added within 1 year of release

constraint_country=df['country'].apply(lambda x: str(x).split(',')).tolist()
df_new_country=pd.DataFrame(constraint_country,index=df['title']).stack()
df_new_country = pd.DataFrame(df_new_country)
df_new_country.reset_index(inplace=True) #to give index to a both title and cast, otherwise will show
df_new_country = df_new_country[['title',0]]
df_new_country.columns = ['title', 'country']
df_new_country_no_nan = df_new_country.replace('nan', np.nan).dropna()

df_country=df.reset_index().merge(df_new_country_no_nan, on='title', how='inner')
df_country_date_added = df_country[df_country["date_added_period"]<1]
df_country_date_added.groupby(["type", "country_y"])['title'].nunique().sort_values(ascending=False).h
```

Out[19]:

type	country_y	
Movie	United States	881
TV Show	United States	566
Movie	India	191
	United Kingdom	129
TV Show	United Kingdom	117
Movie	Canada	86
TV Show	Japan	80
	South Korea	74
Movie	France	72
	Spain	62

Name: title, dtype: int64

In [20]:

```
# Ratings with all movies/tv shows added within 5 yrs of release
df.groupby("rating").filter(lambda x: x["date_added_period"].max() <= 5)
```

Out[20]:

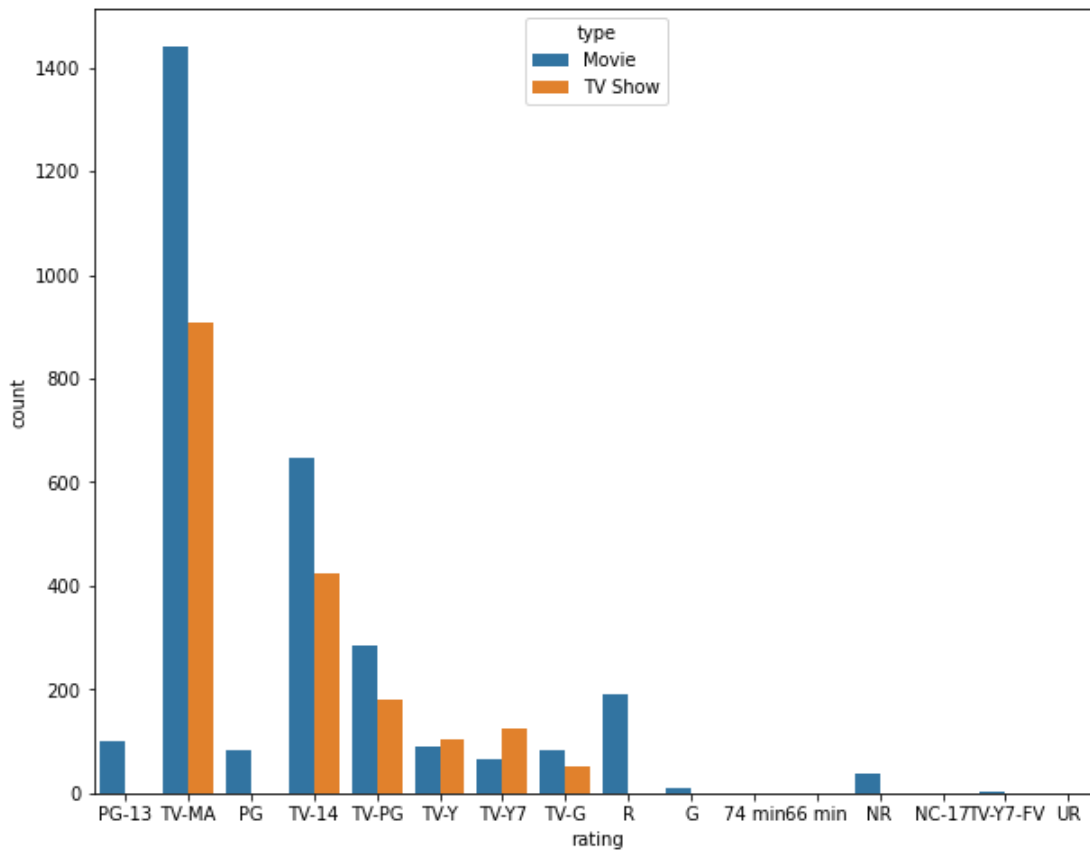
	show_id	type	title	director	cast	country	date_added	release_year	rating	durati
5264	s5265	Movie	Swearnet: The Movie	Warren P. Sonoda	Mike Smith, John Paul Tremblay, Robb Wells, Pa...	Canada	September 24, 2017	2014	NC-17	112 n
5541	s5542	Movie	Louis C.K. 2017	Louis C.K.	Louis C.K.	United States	April 4, 2017	2017	74 min	Ni
5813	s5814	Movie	Louis C.K.: Live at the Comedy Store	Louis C.K.	Louis C.K.	United States	August 15, 2016	2015	66 min	Ni
6338	s6339	Movie	Blue Is the Warmest Color	Abdellatif Kechiche	Léa Seydoux, Adèle Exarchopoulos, Salim Kechio...	France, Belgium, Spain	August 26, 2016	2013	NC-17	180 n
6581	s6582	Movie	Dear Dracula	Chad Van De Keere	Nathan Gamble, Ray Liotta, Emilio Estevez, Ari...	United States	February 18, 2016	2012	TV-Y7-FV	43 n
7263	s7264	Movie	Lady-Like	NaN	Allie Gallerani, Beverly D'Angelo, Olivia Lucc...	United States	January 18, 2019	2018	NC-17	83 n
7290	s7291	Movie	LEGO Ninjago: Masters of Spinjitzu: Day of the...	NaN	NaN	Denmark, China	February 1, 2019	2016	TV-Y7-FV	44 n
7292	s7293	Movie	Leo the Lion	Mario Cambi	Daniel Amerman, John Cygan, Matthew Mercer, Am...	United States, Italy	December 20, 2015	2013	TV-Y7-FV	78 n
7317	s7318	Movie	Little Singham aur Kaal ka Mahajaal	Prakash Satam	NaN	NaN	December 1, 2018	2018	TV-Y7-FV	68 n
7513	s7514	Movie	Motu Patlu: King of Kings	Suhas Kadav	Saurav Chakrabarty, Vinay Pathak	India	July 6, 2017	2016	TV-Y7-FV	109 n
7646	s7647	TV Show	Oh No! It's an Alien Invasion	NaN	Al Mukadam, Dan Chameroy, Seán Cullen, Stacey ...	Canada	May 31, 2015	2014	TV-Y7-FV	Seasc

In [21]:

```
plt.figure(figsize=(10,8))
sns.countplot(x='rating', hue='type', data=df[df["date_added_period"]<=1])
```

Out[21]:

<AxesSubplot:xlabel='rating', ylabel='count'>



In [22]:

```
df_date_added_period_ratings_all = df.groupby("rating").filter(lambda x:x["date_added_period"].max()<=1)
plt.figure(figsize=(10,8))
sns.countplot(x='rating', hue='type', data=df_date_added_period_ratings_all)
```



Duration (runtime and number of seasons)

In [23]:

```
# use apply func
def extract_int(x):
    if type(x) == str:
        x = int(x.split(' ')[0])
    return x

df["duration"] = df["duration"].apply(extract_int)
```

Missing Value Imputation

In [24]:

```
#Unnesting full data
df_final_1 = df.reset_index().merge(df_new_cast.replace('nan',np.nan), on="title", how='inner')
df_final_2 = df_final_1.reset_index().merge(df_new_director.replace('nan',np.nan), on="title", how='inner')
df_final_2 = df_final_2.drop(columns=["level_0"])
df_final_3 = df_final_2.reset_index().merge(df_new_country.replace('nan',np.nan), on="title", how='inner')

df_final = df_final_3.merge(df_new_genre.replace('nan',np.nan), on="title", how='inner')
df_final = df_final.drop(columns=["level_0", "index"])
df_final
```

Out[24]:

	show_id	type	title	director_x	cast_x	country_x	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90.
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2.
2	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2.
3	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2.
4	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2.
...
201986	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111.
201987	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111.
201988	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111.
201989	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111.

show_id	type	title	director_x	cast_x	country_x	date_added	release_year	rating	duration
201990	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Madhav Chanan...	India	March 2, 2019	2015 TV-14	111.

Using listed_in to fill missing values per group

In [25]:
201991 rows × 17 columns

```
import warnings
warnings.filterwarnings('ignore')
df_final['cast_y'] = df_final.groupby('genre')['cast_y'].transform(lambda x: x.fillna(x.mode()[0]))
df_final[df_final["cast_y"].isna()]
```

Out[25]:

show_id	type	title	director_x	cast_x	country_x	date_added	release_year	rating	duration	listed_in	de
---------	------	-------	------------	--------	-----------	------------	--------------	--------	----------	-----------	----

Filling director values using frequently-occured of director in every genre

In [26]:

```
df_final['director_y'] = df_final.groupby('genre')['director_y'].transform(lambda x: x.fillna(x.mode()[0]))
df_final[df_final["director_y"].isna()]
```

Out[26]:

e	title	director_x	cast_x	country_x	date_added	release_year	rating	duration	listed_in	description	date_e
---	-------	------------	--------	-----------	------------	--------------	--------	----------	-----------	-------------	--------

Dropping NAN country values

In [34]:

```
df_final = df_final.dropna()
```

KDEplot

Plotting distribution of top 3 casts (by count) accross the years by kdeplot

In [135]:

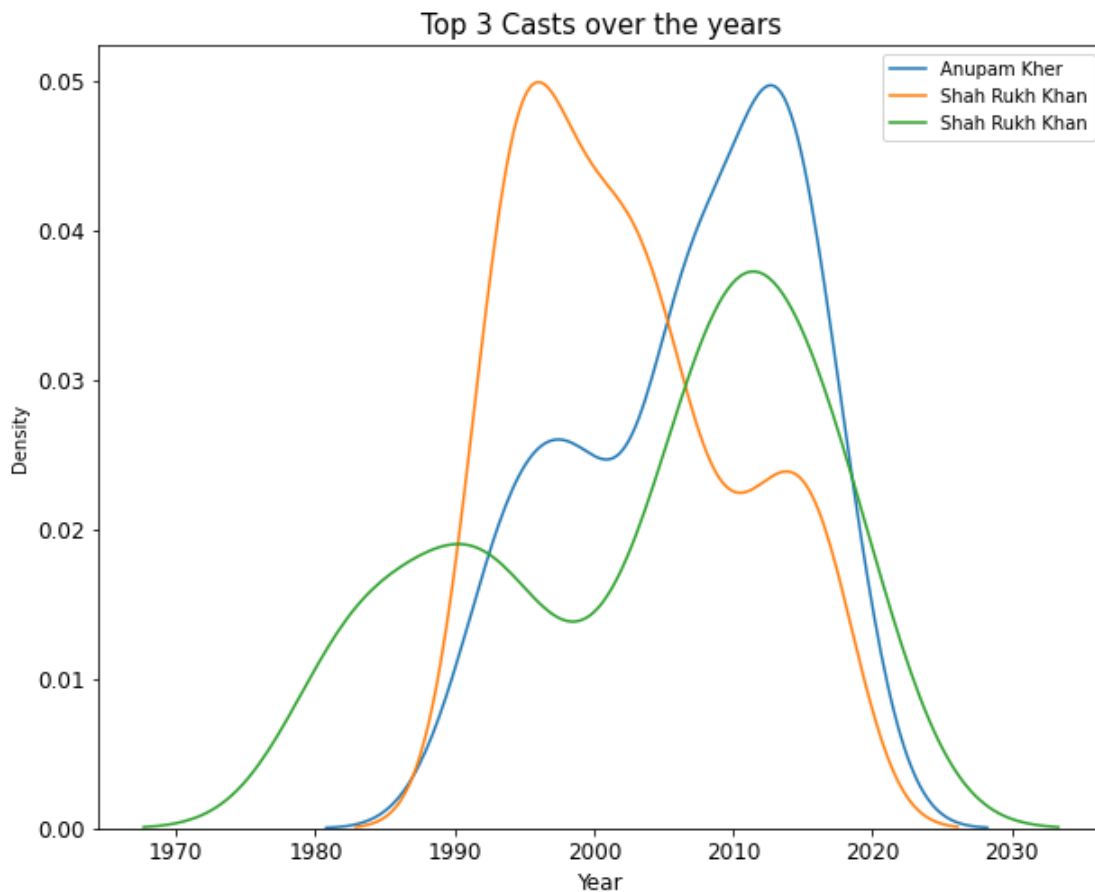
```
import seaborn as sns
df_mask = df_final.groupby('cast_y')['title'].nunique().sort_values(ascending=False).head(3)
#df_final['cast_y'].value_counts()
df_mask_cast = df_mask.keys()
```

In [136]:

```
df_mask_1 = df_final.groupby('genre')['title'].nunique().sort_values(ascending=False).head(3)
df_mask_gen = df_mask_1.keys()
```

In [89]:

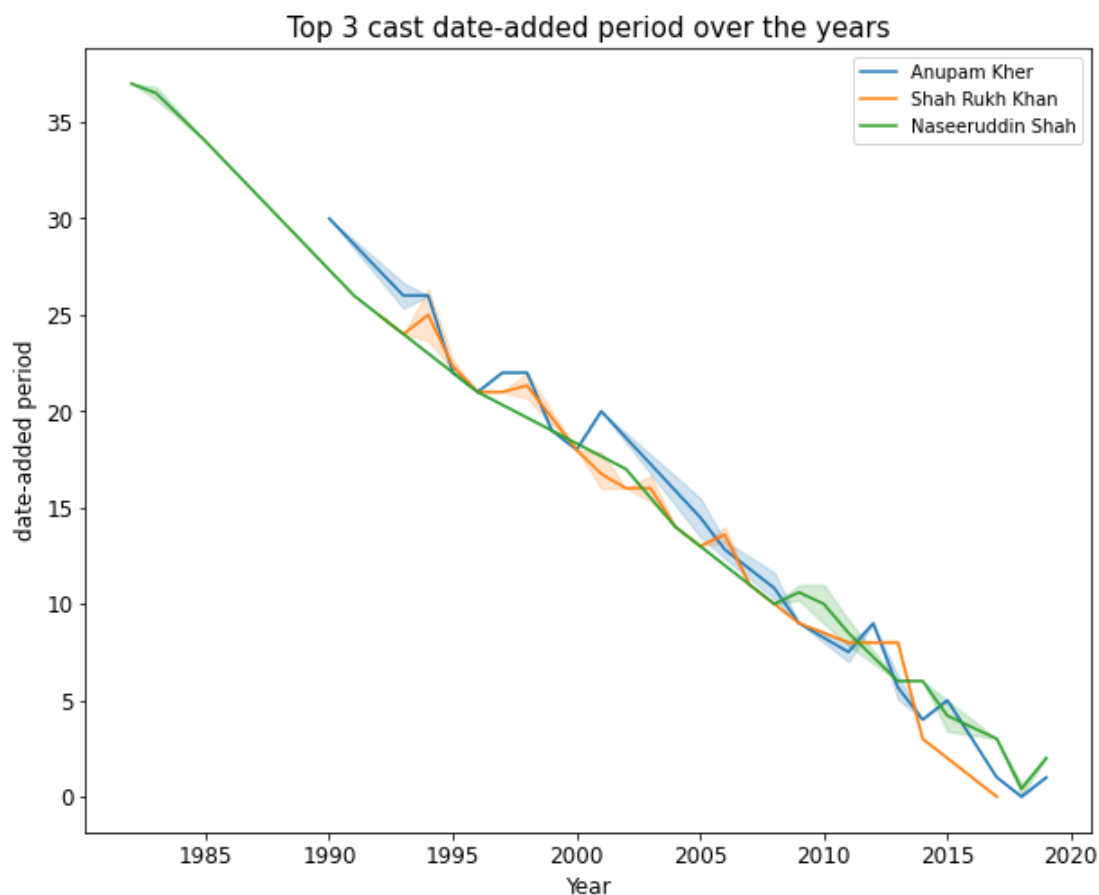
```
plt.figure(figsize=(10,8))
sns.kdeplot(data=df_final.loc[df_final['cast_y']=='Anupam Kher'],x = "release_year", label='Anupam Kher')
sns.kdeplot(data=df_final.loc[df_final['cast_y']=='Shah Rukh Khan'],x = "release_year", label='Shah Rukh Khan')
sns.kdeplot(data=df_final.loc[df_final['cast_y']=='Naseeruddin Shah'],x = "release_year", label='Shah Rukh Khan')
plt.title('Top 3 Casts over the years', fontsize=15)
plt.xlabel('Year', fontsize=12)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.legend()
plt.show()
```



Trendplot

In [98]:

```
# Plotting trend of date-added_period of top 3 cast-director pair (by count) accross the years
plt.figure(figsize=(10,8))
sns.lineplot(x='release_year', y='date_added_period', data=df_final.loc[df_final['cast_y']=='Anupam Kher'])
sns.lineplot(x='release_year', y='date_added_period', data=df_final.loc[df_final['cast_y']=='Shah Rukh Khan'])
sns.lineplot(x='release_year', y='date_added_period', data=df_final.loc[df_final['cast_y']=='Naseeruddin Shah'])
plt.title('Top 3 cast date-added period over the years', fontsize=15)
plt.xlabel('Year', fontsize=12)
plt.ylabel('date-added period', fontsize=12)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.legend()
plt.show()
```



Stacked Countplot

Distribution of genres for top 3 publishers

In [138]:

```
top3_cast = df_final['cast_y'].value_counts().index[:3]
top3_gen = df_final['genre'].value_counts().index[:3]
#top3_dir = df_final['director_y'].value_counts().index[:3]
top3_data = df_final.loc[(df_final["cast_y"].isin(df_mask_cast)) & (df_final["genre"].isin(df_mask_gen))]
top3_data
```


Out[138]:

	show_id	type	title	director_x	cast_x	country_x	date_added	release_year	rating	dui
2491	s115	Movie	Anjaam	Rahul Rawail	Madhuri Dixit, Shah Rukh Khan, Tinnu Anand, Jo...	India	September 2, 2021	1994	TV-14	
2492	s115	Movie	Anjaam	Rahul Rawail	Madhuri Dixit, Shah Rukh Khan, Tinnu Anand, Jo...	India	September 2, 2021	1994	TV-14	
4723	s193	Movie	C Kkompany	Sachin Yardi	Mithun Chakraborty, Tusshar Kapoor, Anupam Khe...	India	August 27, 2021	2008	TV-14	
4724	s193	Movie	C Kkompany	Sachin Yardi	Mithun Chakraborty, Tusshar Kapoor, Anupam Khe...	India	August 27, 2021	2008	TV-14	
5027	s203	Movie	Kyaa Kool Hai Hum	Sangeeth Sivan	Tusshar Kapoor, Riteish Deshmukh, Isha Koppika...	India	August 27, 2021	2005	TV-MA	
...	
199274	s8689	Movie	Waiting	Anu Menon	Naseeruddin Shah, Kalki Koechlin, Rajat Kapoor...	India	May 1, 2018	2015	TV-MA	
201003	s8770	Movie	Y.M.I.: Yeh Mera India	N. Chandra	Anupam Kher, Atul Kulkarni, Aakash Pandey, Mil...	India	March 1, 2018	2008	TV-MA	
201005	s8770	Movie	Y.M.I.: Yeh Mera India	N. Chandra	Anupam Kher, Atul Kulkarni, Aakash Pandey, Mil...	India	March 1, 2018	2008	TV-MA	
201133	s8773	Movie	Yamla Pagla Deewana 2	Sangeeth Sivan	Dharmendra, Sunny Deol, Bobby Deol, Neha Sharm...	India	May 1, 2017	2013	TV-14	
201134	s8773	Movie	Yamla Pagla Deewana 2	Sangeeth Sivan	Dharmendra, Sunny Deol, Bobby Deol, Neha Sharm...	India	May 1, 2017	2013	TV-14	

241 rows × 17 columns

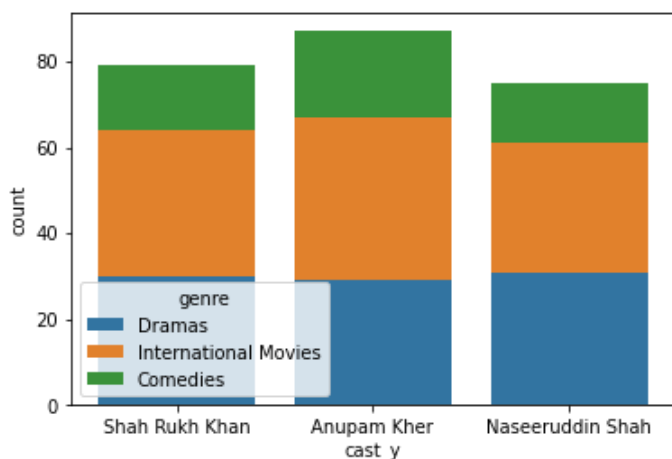


In [139]:

```

ax=sns.countplot(x='cast_y',data=top3_data, hue='genre',dodge=False)
bottoms = {}
for bars in ax.containers:
    for bar in bars:
        x, y = bar.get_xy()
        h = bar.get_height()
        if x in bottoms:
            bar.set_y(bottoms[x])
            bottoms[x] += h
        else:
            bottoms[x] = h
ax.relim() # the plot limits need to be updated with the moved bars
ax.autoscale()
plt.show()

```



Code Explanation:

- We first use `dodge=False` parameter to not form the a dodged countplot
- Now, simply using `dodge=False` will make the position of each bar constant, but won't adjust the heights.
 - The bars will end up overlapping, hence we need to manually update the heights for for the bars
- We use a `bottoms` dictionary, and for each genre, update the bar's height with the data
- After every updation, we set the height of bar using `bar.set_y`
- Finally we use `ax.relim()` and `ax.autoscale()` to update and scale the bars at the end as per the plot

Do print the data in the bottoms dictionary to get a cleared idea of how this is working!

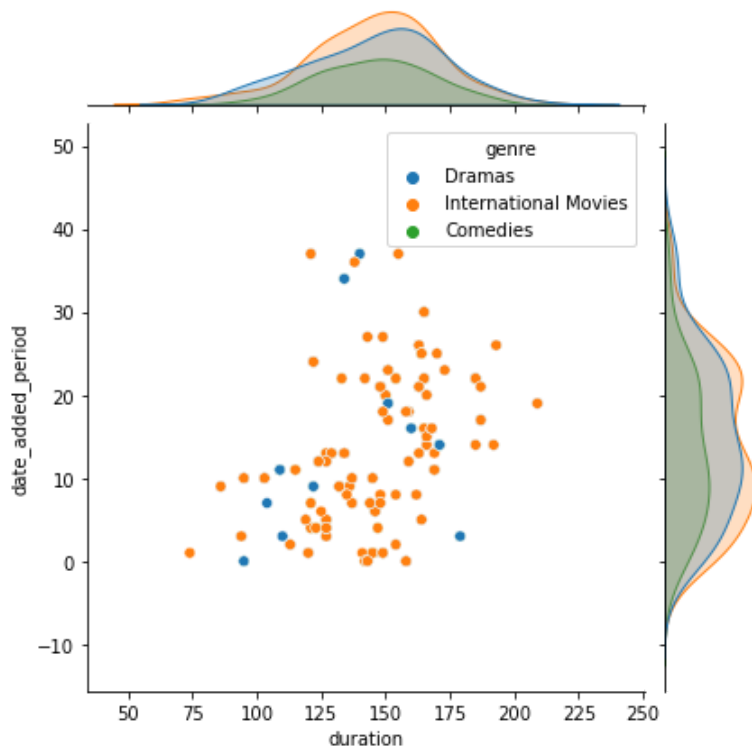
Jointplot between duration and period of date added

In [143]:

```
sns.jointplot(data=top3_data,x="duration",y="date_added_period", hue="genre")
```

Out[143]:

<seaborn.axisgrid.JointGrid at 0x2d94a3d39d0>



In [137]:

```
#df_final['cast_y'].value_counts().index[:3]
```

Out[137]:

	show_id	type	title	director_x	cast_x	country_x	date_added	release_year	rating	dui
	2491	s115	Movie	Anjaam	Rahul Rawail	Madhuri Dixit, Shah Rukh Khan, Tinnu Anand, Jo...	India	September 2, 2021	1994	TV-14
	2492	s115	Movie	Anjaam	Rahul Rawail	Madhuri Dixit, Shah Rukh Khan, Tinnu Anand, Jo...	India	September 2, 2021	1994	TV-14
	2493	s115	Movie	Anjaam	Rahul Rawail	Madhuri Dixit, Shah Rukh Khan, Tinnu Anand, Jo...	India	September 2, 2021	1994	TV-14
	4722	s193	Movie	C Kkompany	Sachin Yardi	Mithun Chakraborty, Tusshar Kapoor, Anupam Khe...	India	August 27, 2021	2008	TV-14
	4723	s193	Movie	C Kkompany	Sachin Yardi	Mithun Chakraborty, Tusshar Kapoor, Anupam Khe...	India	August 27, 2021	2008	TV-14

	201004	s8770	Movie	Y.M.I.: Yeh Mera India	N. Chandra	Anupam Kher, Atul Kulkarni, Aakash Pandey, Mil...	India	March 1, 2018	2008	TV-MA
	201005	s8770	Movie	Y.M.I.: Yeh Mera India	N. Chandra	Anupam Kher, Atul Kulkarni, Aakash Pandey, Mil...	India	March 1, 2018	2008	TV-MA
	201132	s8773	Movie	Yamla Pagla Deewana 2	Sangeeth Sivan	Dharmendra, Sunny Deol, Bobby Deol, Neha Sharm...	India	May 1, 2017	2013	TV-14
	201133	s8773	Movie	Yamla Pagla Deewana 2	Sangeeth Sivan	Dharmendra, Sunny Deol, Bobby Deol, Neha Sharm...	India	May 1, 2017	2013	TV-14
	201134	s8773	Movie	Yamla Pagla Deewana 2	Sangeeth Sivan	Dharmendra, Sunny Deol, Bobby Deol, Neha Sharm...	India	May 1, 2017	2013	TV-14

328 rows × 17 columns



End of my analysis part

In [29]:

```
df_new=df_new[['title',0]]  
  
df_new
```

```
-----  
NameError                                Traceback (most recent call last)  
<ipython-input-29-4ebf6d1a2380> in <module>  
----> 1 df_new=df_new[['title',0]]  
      2  
      3 df_new  
  
NameError: name 'df_new' is not defined
```

In [30]:

```
df_new.columns=['title','cast']  
  
df_new[df_new['cast']=='Rupa Bhimani']
```

```
-----  
NameError                                Traceback (most recent call last)  
<ipython-input-30-7620cc278871> in <module>  
----> 1 df_new.columns=['title','cast']  
      2  
      3 df_new[df_new['cast']=='Rupa Bhimani']  
  
NameError: name 'df_new' is not defined
```

In [31]:

```
df_new
```

```
-----  
NameError                                Traceback (most recent call last)  
<ipython-input-31-94dbd581444b> in <module>  
----> 1 df_new  
  
NameError: name 'df_new' is not defined
```

In [32]:

```
df_new.columns=['title','cast']
```

```
-----  
NameError                                Traceback (most recent call last)  
<ipython-input-32-a7a7b5e2b22b> in <module>  
----> 1 df_new.columns=['title','cast']  
  
NameError: name 'df_new' is not defined
```

In [23]:

```
df_new[df_new['cast']=='Rupa Bhimani']
```

Out[23]:

	title	cast
313	Chhota Bheem	Rupa Bhimani
708	Mighty Raju	Rupa Bhimani
3373	Chhota Bheem - Dinosaur World	Rupa Bhimani
3379	Chhota Bheem - Neeli Pahaadi	Rupa Bhimani
3386	Chhota Bheem & Ganesh	Rupa Bhimani
3393	Chhota Bheem & Krishna: Mayanagari	Rupa Bhimani
3400	Chhota Bheem & Krishna: Pataliputra- City of t...	Rupa Bhimani
3407	Chhota Bheem And The Broken Amulet	Rupa Bhimani
3414	Chhota Bheem And The Crown of Valhalla	Rupa Bhimani
3421	Chhota Bheem and the Incan Adventure	Rupa Bhimani
3428	Chhota Bheem and The ShiNobi Secret	Rupa Bhimani
3435	Chhota Bheem Aur Hanuman	Rupa Bhimani
3442	Chhota Bheem aur Krishna	Rupa Bhimani
3449	Chhota Bheem aur Krishna vs Zimbara	Rupa Bhimani
3455	Chhota Bheem in African Safari	Rupa Bhimani
3463	Chhota Bheem Ka Roosi Romanch	Rupa Bhimani
3471	Chhota Bheem: Bheem vs Aliens	Rupa Bhimani
3478	Chhota Bheem: Dholakpur to Kathmandu	Rupa Bhimani
3485	Chhota Bheem: Dus Pe Dus	Rupa Bhimani
3492	Chhota Bheem: Journey to Petra	Rupa Bhimani
3498	Chhota Bheem: Master of Shaolin	Rupa Bhimani
3504	Chhota Bheem: The Rise of Kirmada	Rupa Bhimani
45767	Antariksha Ke Rakhwale	Rupa Bhimani
47895	Chhota Bheem Aur Kaala Yodha	Rupa Bhimani
47904	Chhota Bheem Ka Romani Adventure	Rupa Bhimani
47913	Chhota Bheem Ka Troll Se Takkar	Rupa Bhimani
47920	Chhota Bheem Kungfu Dhamaka	Rupa Bhimani
53418	Krishna Balram	Rupa Bhimani
58346	Samudri Lootere	Rupa Bhimani
59835	Super Bheem	Rupa Bhimani
59841	Super Bheem Bana Vajraveer	Rupa Bhimani

In [60]:

```
df_new[df_new.cast.str.contains('Rupa')]
```

Out[60]:

	title	cast
313	Chhota Bheem	Rupa Bhimani
708	Mighty Raju	Rupa Bhimani
2108	Pahuna	Rupa Tamang
3373	Chhota Bheem - Dinosaur World	Rupa Bhimani
3379	Chhota Bheem - Neeli Pahaadi	Rupa Bhimani
3386	Chhota Bheem & Ganesh	Rupa Bhimani
3393	Chhota Bheem & Krishna: Mayanagari	Rupa Bhimani
3400	Chhota Bheem & Krishna: Pataliputra- City of t...	Rupa Bhimani
3407	Chhota Bheem And The Broken Amulet	Rupa Bhimani
3414	Chhota Bheem And The Crown of Valhalla	Rupa Bhimani
3421	Chhota Bheem and the Incan Adventure	Rupa Bhimani
3428	Chhota Bheem and The ShiNobi Secret	Rupa Bhimani
3435	Chhota Bheem Aur Hanuman	Rupa Bhimani
3442	Chhota Bheem aur Krishna	Rupa Bhimani
3449	Chhota Bheem aur Krishna vs Zimbara	Rupa Bhimani
3455	Chhota Bheem in African Safari	Rupa Bhimani
3463	Chhota Bheem Ka Roosi Romanch	Rupa Bhimani
3471	Chhota Bheem: Bheem vs Aliens	Rupa Bhimani
3478	Chhota Bheem: Dholakpur to Kathmandu	Rupa Bhimani
3485	Chhota Bheem: Dus Pe Dus	Rupa Bhimani
3492	Chhota Bheem: Journey to Petra	Rupa Bhimani
3498	Chhota Bheem: Master of Shaolin	Rupa Bhimani
3504	Chhota Bheem: The Rise of Kirmada	Rupa Bhimani
34026	Barfi!	Rupa Ganguly
35016	Haani	Rupan Bal
43471	X: Past Is Present	Pooja Ruparel
45767	Antariksha Ke Rakhwale	Rupa Bhimani
47895	Chhota Bheem Aur Kaala Yodha	Rupa Bhimani
47904	Chhota Bheem Ka Romani Adventure	Rupa Bhimani
47913	Chhota Bheem Ka Troll Se Takkar	Rupa Bhimani
47920	Chhota Bheem Kungfu Dhamaka	Rupa Bhimani
53418	Krishna Balram	Rupa Bhimani
58346	Samudri Lootere	Rupa Bhimani
59835	Super Bheem	Rupa Bhimani
59841	Super Bheem Bana Vajraveer	Rupa Bhimani

In [61]:

```
df_new.cast.value_counts()
```

Out[61]:

```
nan                825
Anupam Kher        43
Shah Rukh Khan     35
Julie Teiwani      33
Naseeruddin Shah   32
...
Melanie Straub     1
Gabriela Maria Schmeide 1
Helena Zengel      1
Daniel Valenzuela  1
Chittaranjan Tripathy 1
Name: cast, Length: 36440, dtype: int64
```

In [24]:

```
df.cast.value_counts()
```

Out[24]:

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

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

In [62]:

```
df_cast=df.reset_index().merge(df_new, on='title', how='inner')
```


In [64]:

```
df_cast[df_cast['cast_y']=='Anupam Kher']
```

1718	204	s205	Movie	Kyaa Super Kool Hain Hum	Sachin Yardi	Tusshar Kapoor, Riteish Deshmukh, Anupam Kher,...	India	August 27, 2021	2012	TV-MA	136 min
1728	205	s206	Movie	Kyo Kii... Main Jhuth Nahin Bolta	David Dhawan	Govinda, Sushmita Sen, Rambha, Anupam Kher, Sa...	India	August 27, 2021	2001	TV-14	150 min
5985	702	s703	Movie	Silver Linings Playbook	David O. Russell	Bradley Cooper, Jennifer Lawrence, Robert De N...	United States	June 17, 2021	2012	R	122 min
Aamir Khan, Madhuri											

In [25]:

```
x=pd.DataFrame(df_cast.groupby('cast_y')['title'].count())
```

In [26]:

```
x.reset_index(inplace=True)
```

In [27]:

```
x[x['cast_y']=='Anupam Kher']
```

Out[27]:

	cast_y	title
34330	Anupam Kher	4

In [28]:

```
df_cast[df_cast['cast_y']=='Anupam Kher']
```

Out[28]:

	index	show_id	type	title	director	cast_x	country	date_added	release_year	rating
20434	2564	s2565	Movie	One Day: Justice Delivered	Ashok Nanda	Anupam Kher, Esha Gupta, Kumud Mishra, Kashyap...	India	May 3, 2020	2019	TV-14
44961	6063	s6064	Movie	A Wednesday	Neeraj Pandey	Anupam Kher, Naseeruddin Shah, Jimmy Shergill,...	India	November 1, 2018	2008	TV- MA
53074	7186	s7187	Movie	Khosla Ka Ghosla	Dibakar Banerjee	Anupam Kher, Boman Irani, Parvin Dabas, Tara S...	India	August 2, 2018	2006	TV-14
64642	8769	s8770	Movie	Y.M.I.: Yeh Mera India	N. Chandra	Anupam Kher, Atul Kulkarni, Aakash Pandey, Mil...	India	March 1, 2018	2008	TV- MA

In [29]:

```
df_cast[df_cast['title']=='A Wednesday']
```

Out[29]:

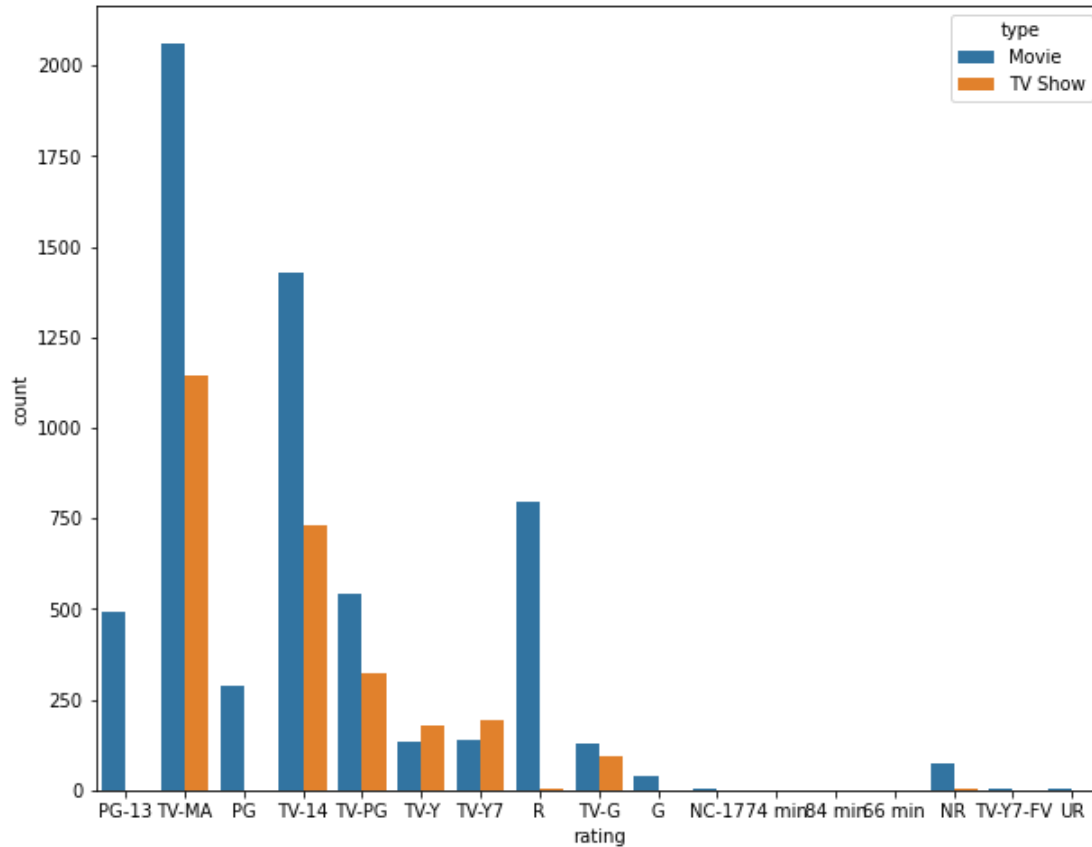
	index	show_id	type	title	director	cast_x	country	date_added	release_year	rating
44961	6063	s6064	Movie	A Wednesday	Neeraj Pandey	Anupam Kher, Naseeruddin Shah, Jimmy Shergill,...	India	November 1, 2018	2008	TV-MA
44962	6063	s6064	Movie	A Wednesday	Neeraj Pandey	Anupam Kher, Naseeruddin Shah, Jimmy Shergill,...	India	November 1, 2018	2008	TV-MA
44963	6063	s6064	Movie	A Wednesday	Neeraj Pandey	Anupam Kher, Naseeruddin Shah, Jimmy Shergill,...	India	November 1, 2018	2008	TV-MA
44964	6063	s6064	Movie	A Wednesday	Neeraj Pandey	Anupam Kher, Naseeruddin Shah, Jimmy Shergill,...	India	November 1, 2018	2008	TV-MA
44965	6063	s6064	Movie	A Wednesday	Neeraj Pandey	Anupam Kher, Naseeruddin Shah, Jimmy Shergill,...	India	November 1, 2018	2008	TV-MA
44966	6063	s6064	Movie	A Wednesday	Neeraj Pandey	Anupam Kher, Naseeruddin Shah, Jimmy Shergill,...	India	November 1, 2018	2008	TV-MA
44967	6063	s6064	Movie	A Wednesday	Neeraj Pandey	Anupam Kher, Naseeruddin Shah, Jimmy Shergill,...	India	November 1, 2018	2008	TV-MA
44968	6063	s6064	Movie	A Wednesday	Neeraj Pandey	Anupam Kher, Naseeruddin Shah, Jimmy Shergill,...	India	November 1, 2018	2008	TV-MA

In [96]:

```
import seaborn as sns
import matplotlib.pyplot as plt
plt.figure(figsize=(10,8))
sns.countplot(x='rating', hue='type', data=df)
```

Out[96]:

<AxesSubplot:xlabel='rating', ylabel='count'>

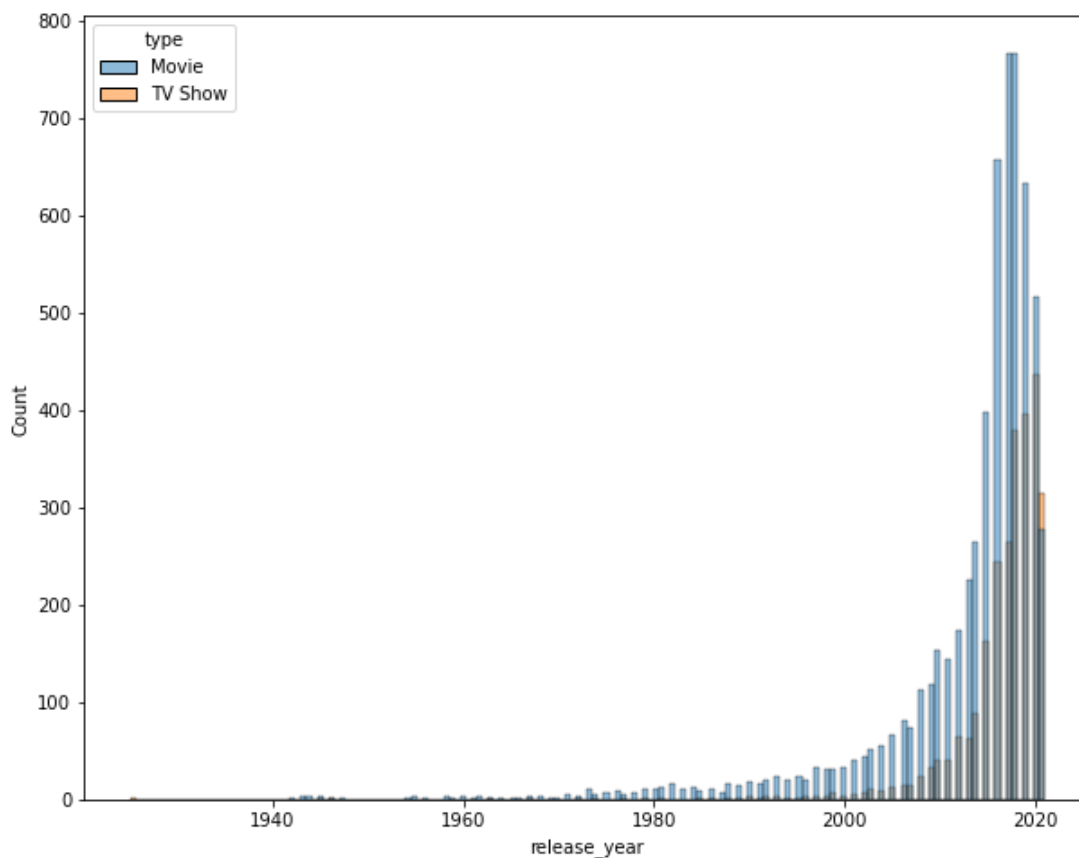


In [31]:

```
plt.figure(figsize=(10,8))
sns.histplot(x='release_year',hue='type',data=df)
```

Out[31]:

<AxesSubplot:xlabel='release_year', ylabel='Count'>



In [32]:

```
df['date_add']=pd.to_datetime(df['date_added'], errors='coerce')
```

In [33]:

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 13 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
12  date_add        8797 non-null   datetime64[ns]
dtypes: datetime64[ns](1), int64(1), object(11)
memory usage: 894.6+ KB
```

In [34]:

```
df.head()
```

Out[34]:

	show_id	type		title	director	cast	country	date_added	release_year	rating	duration	
0	s1	Movie		Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Docu
1	s2	TV Show		Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	Int TV S Dr
2	s3	TV Show		Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Int TV S
3	s4	TV Show		Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Dr f
4	s5	TV Show		Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	Int T Roi Shc

In [35]:

```
df['month']=df['date_add'].dt.month.fillna(-1)
```

In [36]:

```
df['month']=df['month'].astype('int64')
```

In [37]:

```
df.head()
```

Out[37]:

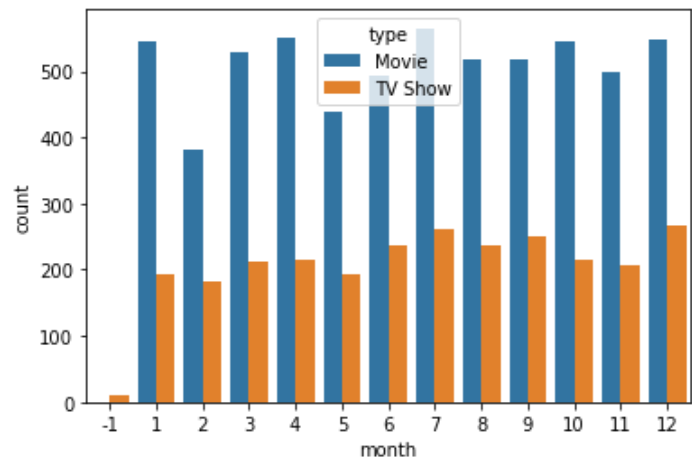
	show_id	type		title	director	cast	country	date_added	release_year	rating	duration	
0	s1	Movie		Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Docu
1	s2	TV Show		Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	Int TV S Dr
2	s3	TV Show		Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Int TV S
3	s4	TV Show		Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Dr f
4	s5	TV Show		Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	Int T Roi Shc

In [38]:

```
sns.countplot(x='month',hue='type', data=df)
```

Out[38]:

<AxesSubplot:xlabel='month', ylabel='count'>



In []:

```
#issue with ratings
```

In [25]:

```
df.rating.value_counts()
```

Out[25]:

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 []: